次世代シーケンサーデータの解析手法
第8回アセンブリ後の解析：ウェブ資料

谷澤 靖洋、神沼英里*、中村保一、遠野雅徳 、
寺田朋子、清水謙多郎、門田 幸二*

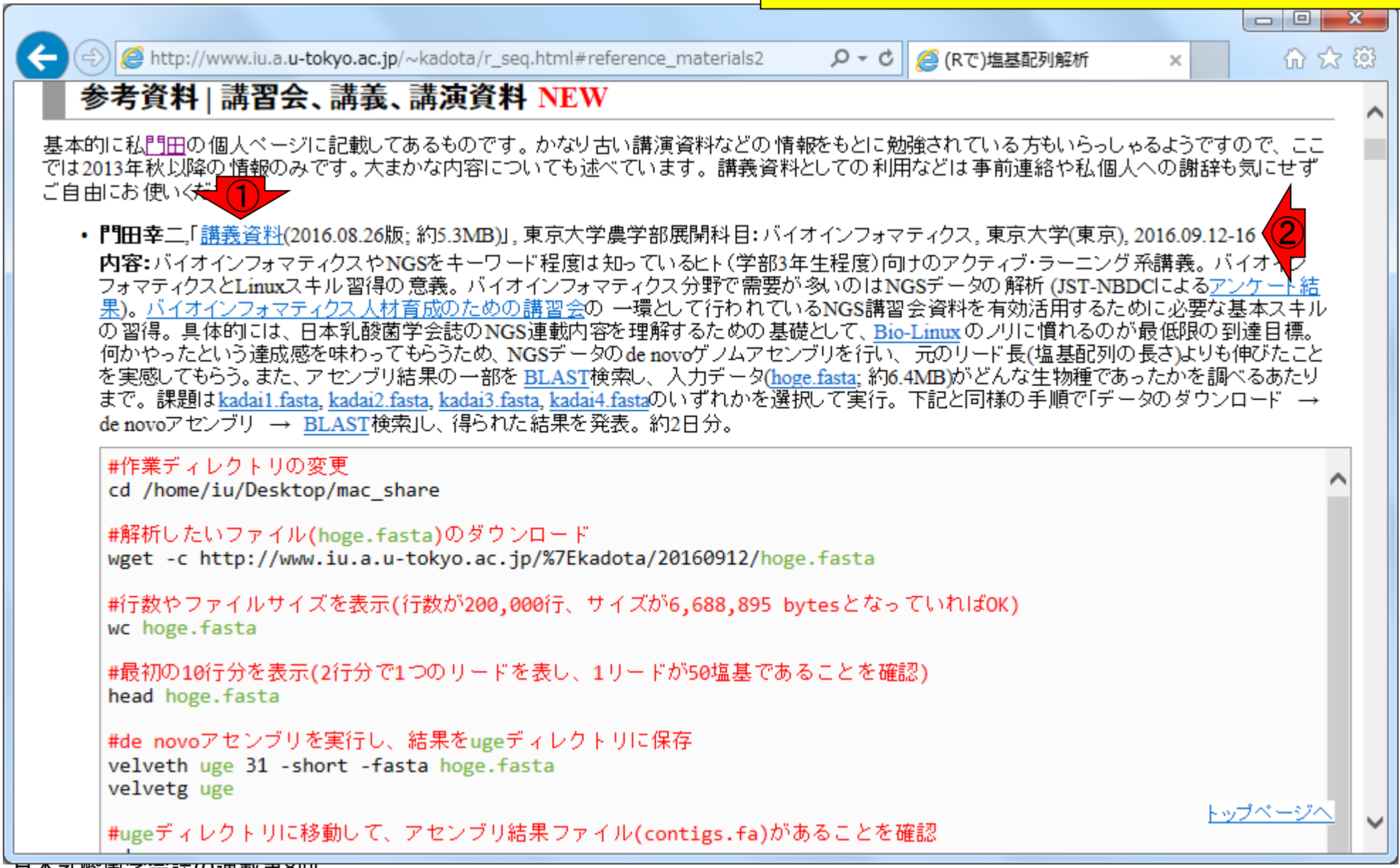
*東京大学・大学院農学生命科学研究科
kadota@bi.a.u-tokyo.ac.jp
http://www.iu.a.u-tokyo.ac.jp/~kadota/

W1：初心者向け教材

①は責任著者（門田）が公開している講習会、講義、講演資料の項目

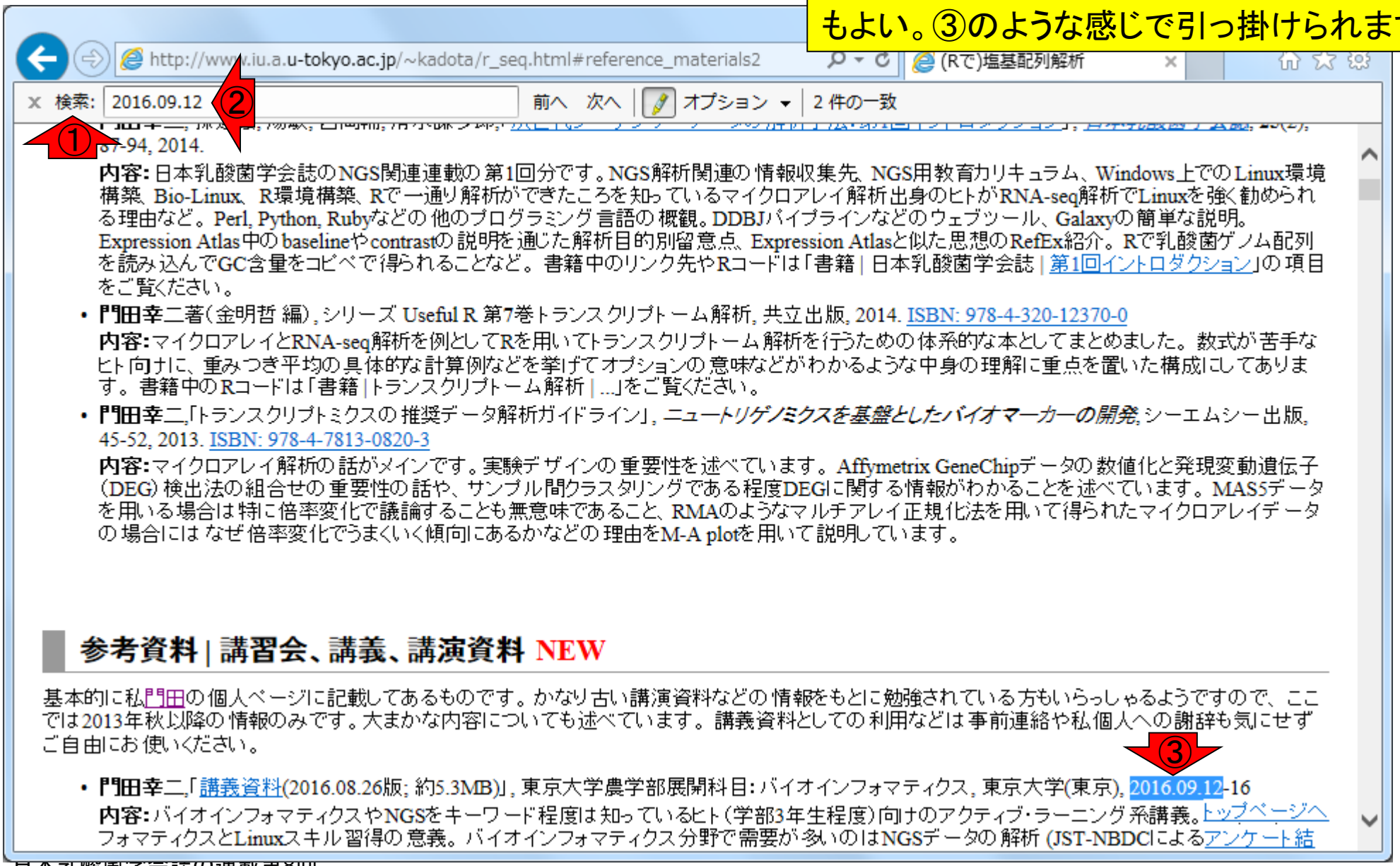
W1：初心者向け教材

新しいものから順にリストアップされている。①今回の目的物は、②2016年9月12-16日の講義資料なので、たまたま一番上に見えているだけ



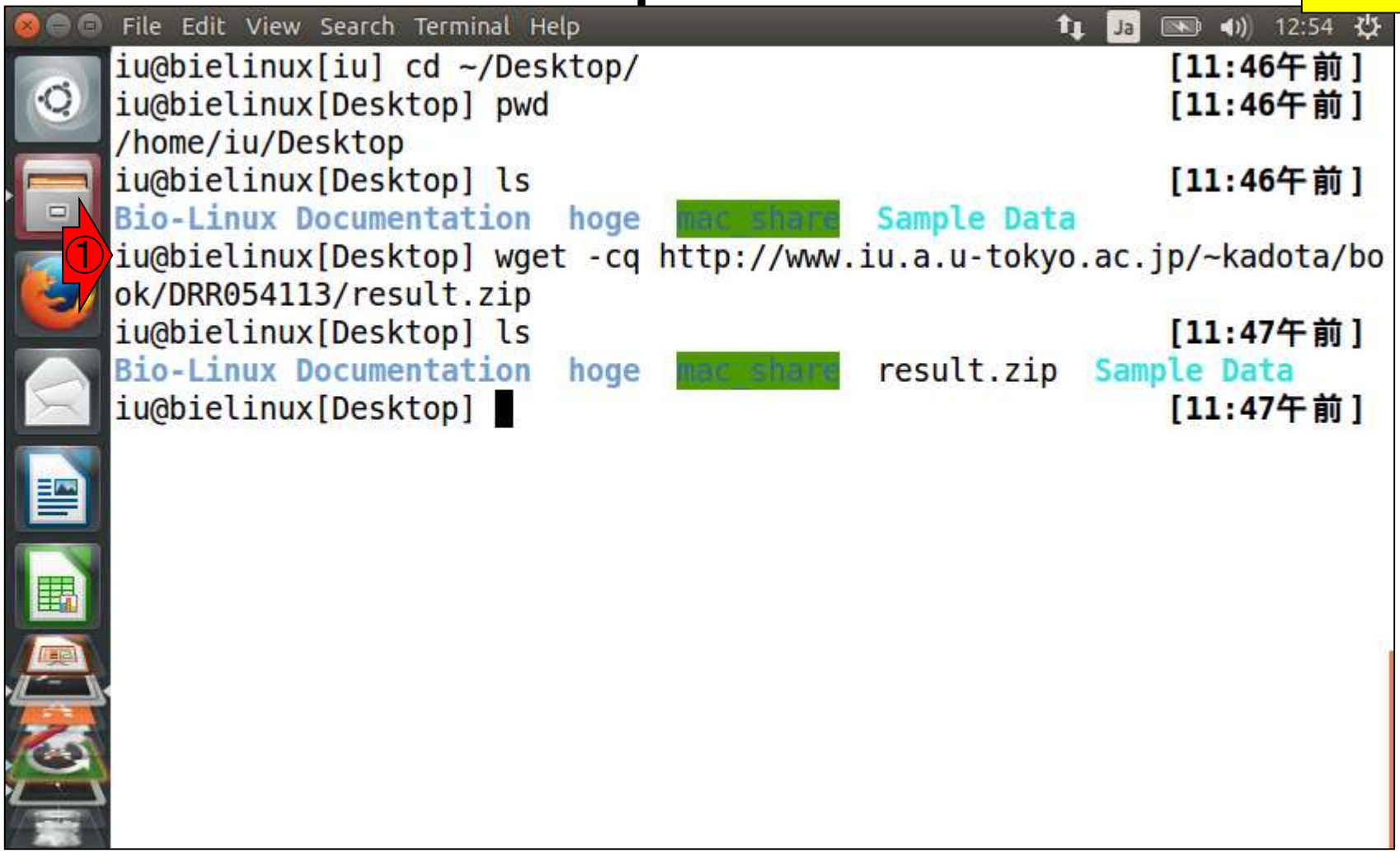
W1：初心者向け教材

Windowsの場合は「CTRL + F」を押して、①ページ内検索で、目的とする講義資料の②開催日時（この場合2016.09.12）で検索してもよい。③のような感じで引っ掛けられます



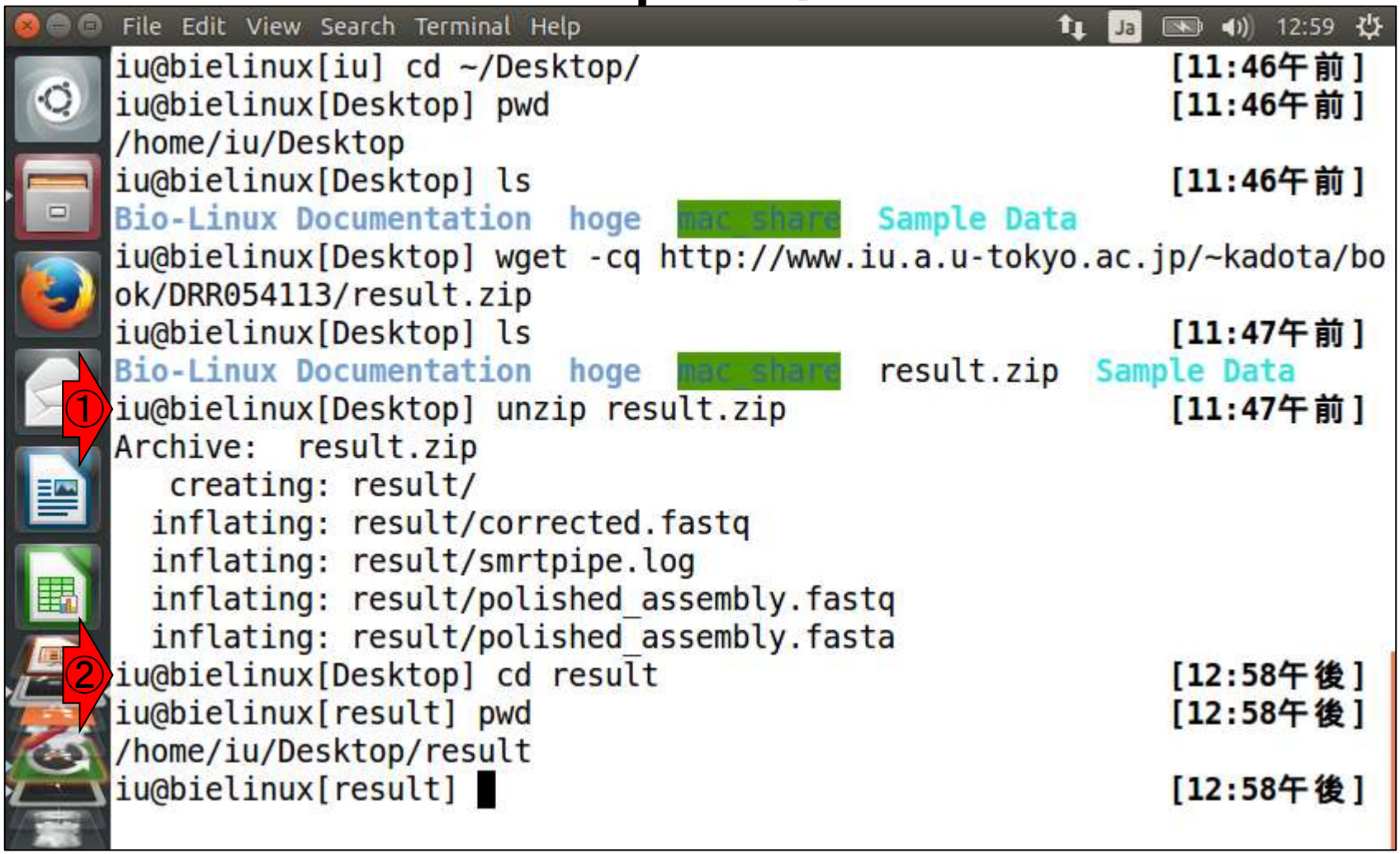
W2-1：result.zipダウンロード

①HGAP実行結果ファイル(result.zip)のダウンロード。第6回W9-3



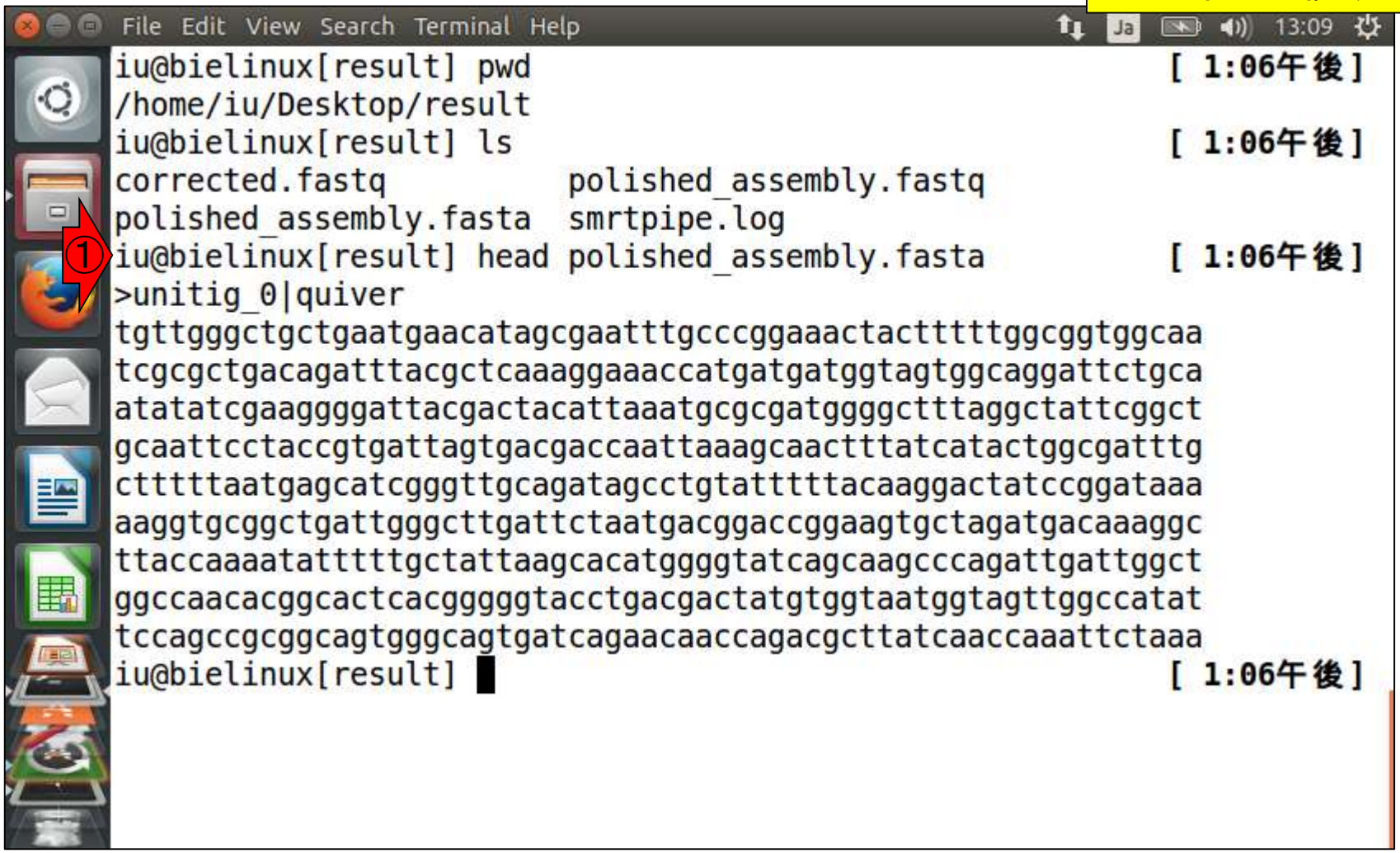
W2-2：result.zipの解凍

①unzipで解凍（第6回W9-4）。②resultディレクトリに移動



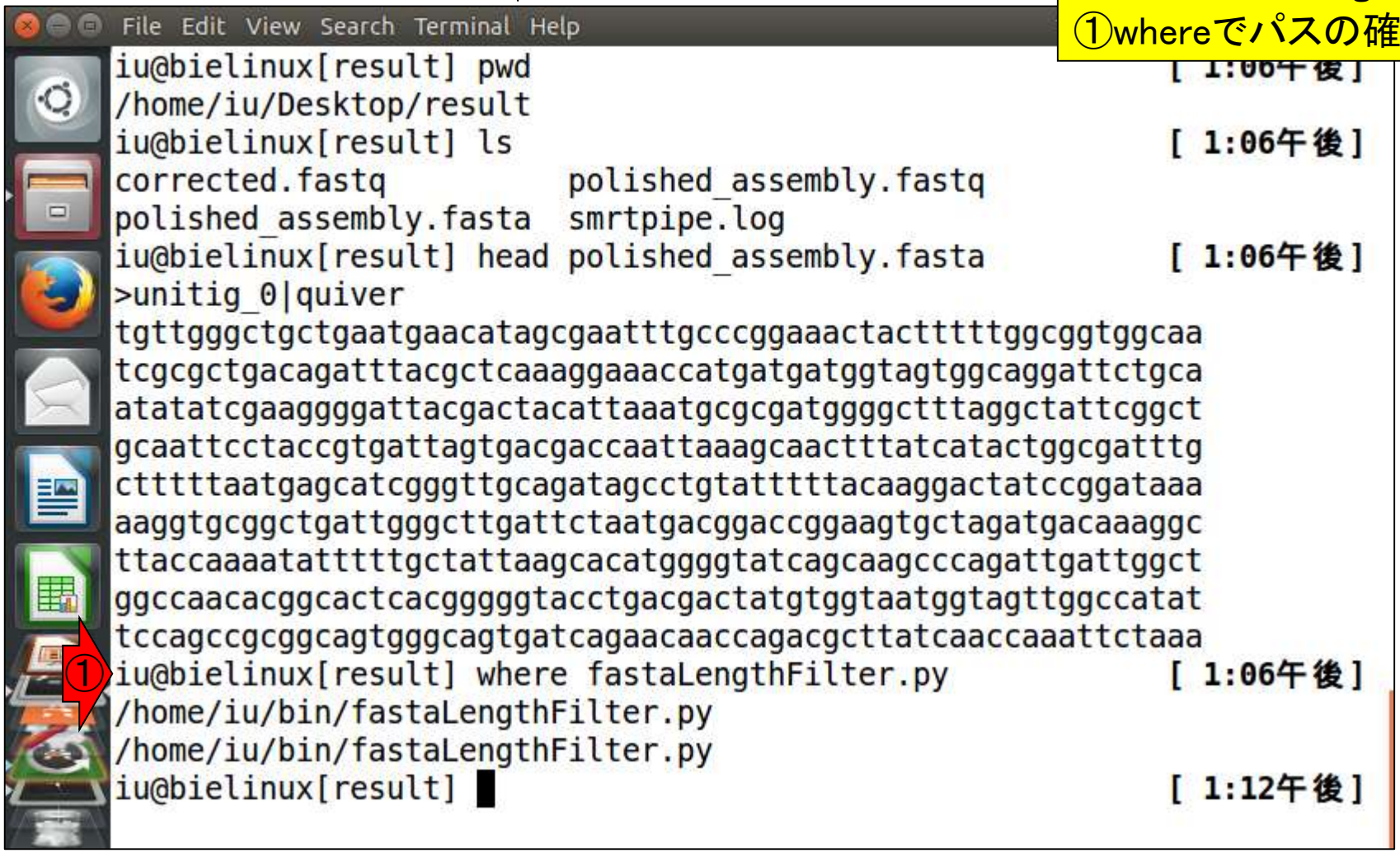
W2-3：FASTAファイル

①アセンブリ結果のFASTAファイル(polished_assembly.fasta)は、塩基配列部分が複数行に分かれている



W2-4：配列長でソート

配列長（降順）でソートし、コンティグの塩基配列部分を1行にすることを目的としてfastaLengthFilter.pyを利用。
①whereでパスの確認。第6回W12

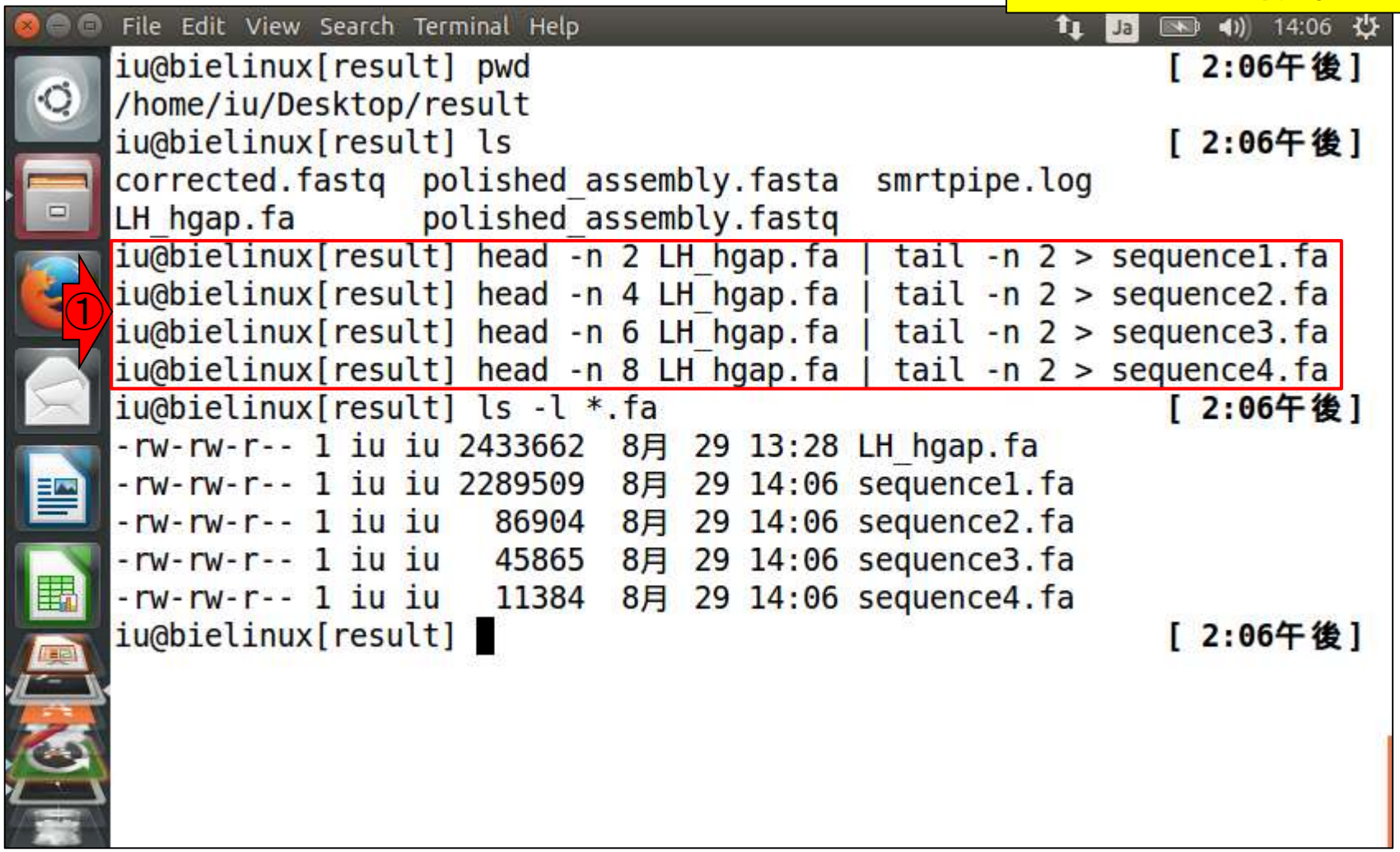


W2-5：実行

①polished_assembly.fastaを入力として、0塩基以上の配列を抽出した結果をLH_hgap.faというファイル名で保存するfastaLengthFilter.pyを実行。オリジナルは4コンティグのFASTAファイルなので、②wc実行結果が③8になるのは妥当

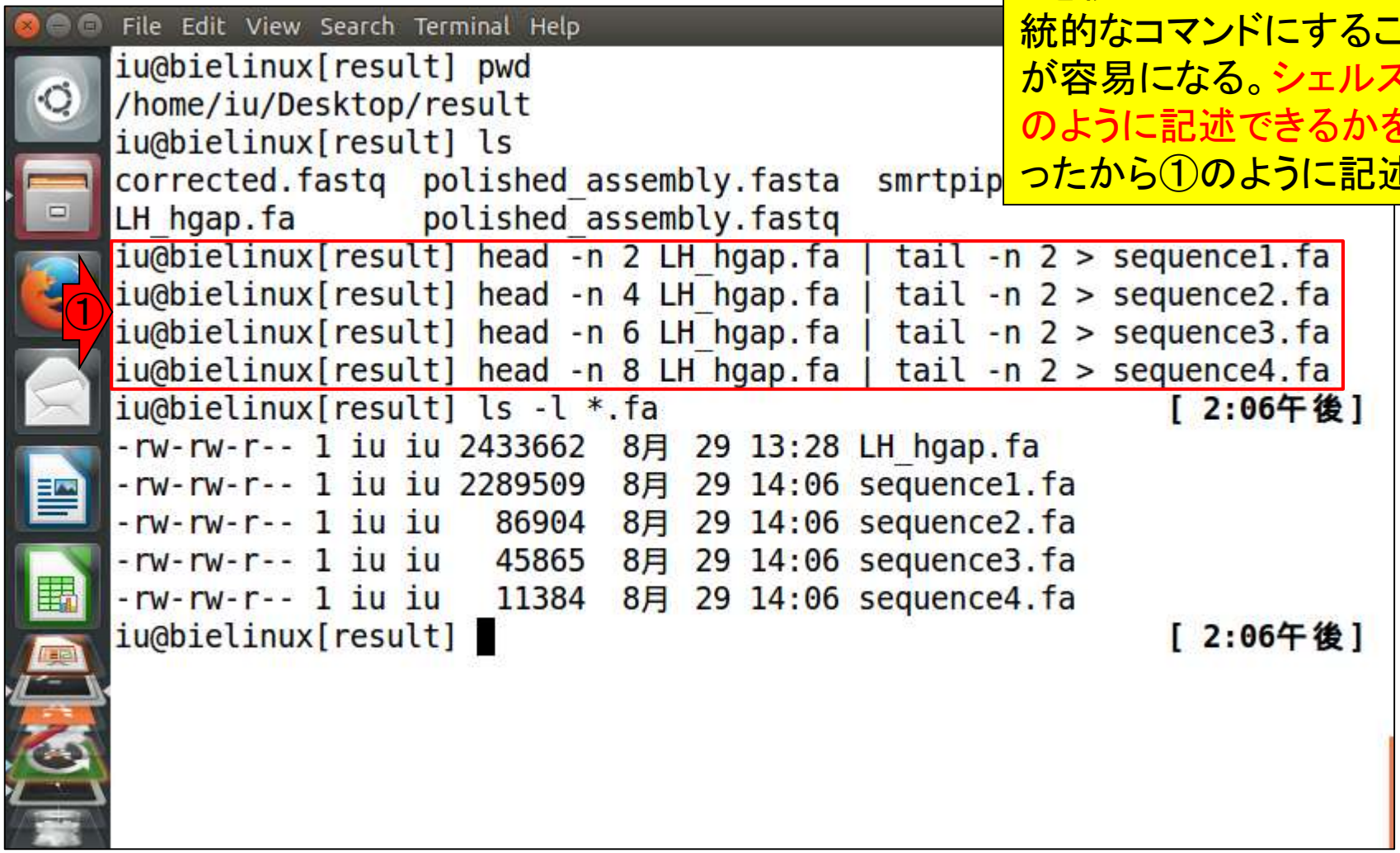
W2-6：ファイル分割

①headとtailコマンドを組み合わせて、multi-FASTAファイルからsingle-FASTAファイルに分割。第7回W10-3と10-4



W3-1：シェルスクリプト

第7回W10-3やW10-4とは異なり、①sequence[0-9].fa作成のためのスクリプトを統一的に記述している。このような系統的なコマンドにすることで、自動処理が容易になる。シェルスクリプトだと、どのように記述できるかを示す（示したかったから①のように記述しているのです）



W3-2：参考資料

参考資料は、例えば①平成27年度NGSハンズオン講習会の②7/24の講義資料のスライド58以降です

平成27年度NGSハンズオン講習会 ①

平成27年度は、平成26年度の実績を踏まえ、講義内容の改善等を行い、ハンズオンに特化した、より効果的なNGS講習会を開催しました。

H27年度概要

H27年度講義日程・参考資料

H26年度講習会の情報についてはこちらをご覧ください。

H27年度実施報告書・講義資料・動画等

● 講習会実施報告書（PDF：2.17MB）および受講者アンケート集計結果（データ集）（PDF：662KB）

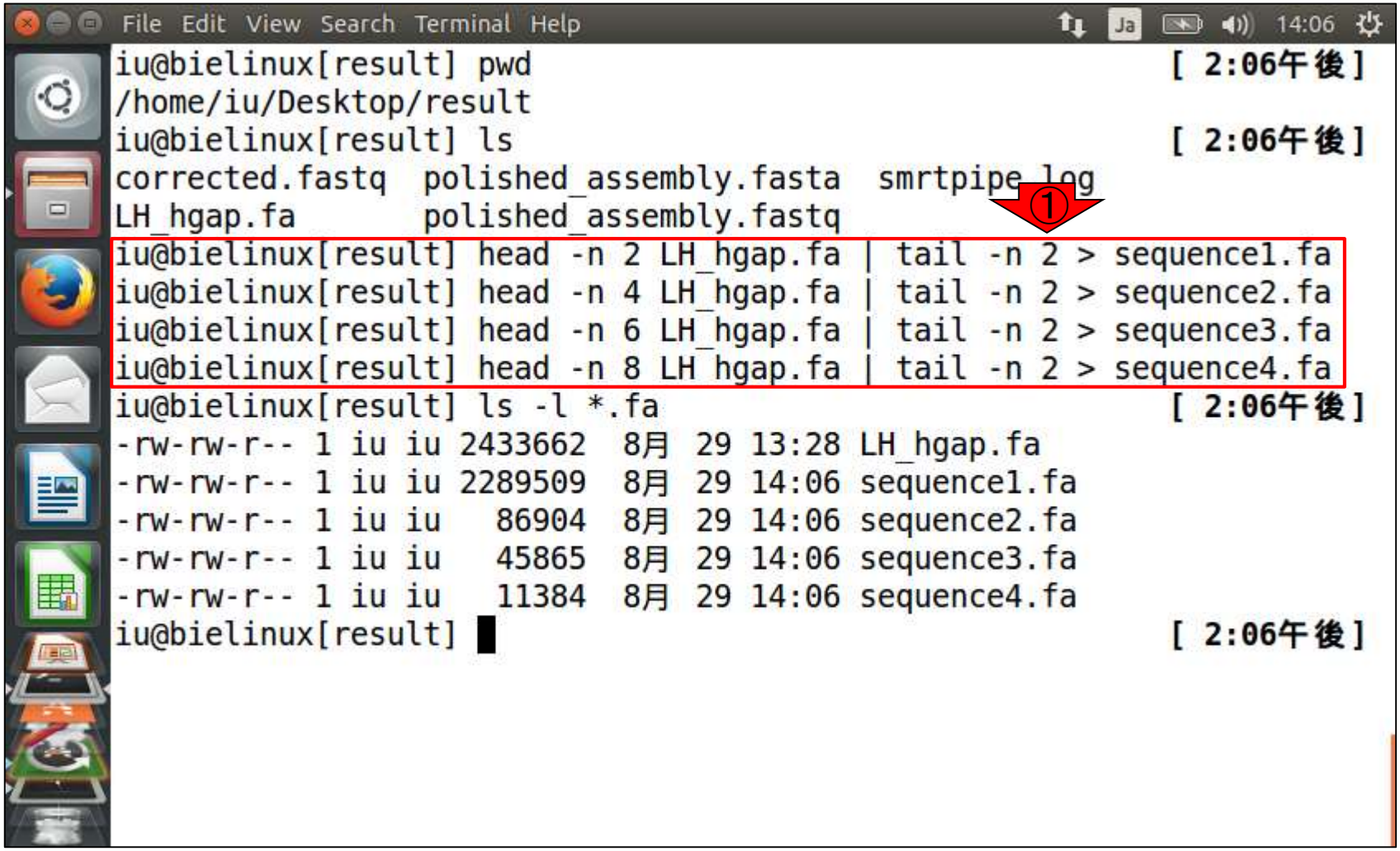
● 講義資料・動画　*講義資料一覧のファイル名をクリックすると資料ファイル（PDF等）がダウンロードできます。

実施日	実施時間	大項目	項目	レベル	習得技術	担当講師(敬称略)	講義資料・動画(統合TV)
7月22日（水）	10:30-12:00	PC環境の構築	Bio-Linux8とRのインストール状況確認		・Linux導入 ・R導入 ・NGS解析に必要な環境構築技術	門田 幸二（東京大学） 寺田 透（東京大学）	事前予習資料一覧（PDF:52KB） 講義資料一覧（PDF:108KB）
	13:15-14:45						
	15:00-16:30						
	16:45-18:15						
7月23日（木）	10:30-12:00	UNIX/Linuxとスクリプト言語	Linux基礎	初級	UNIXの基礎の理解	門田 幸二（東京大学）	講義資料一覧（PDF:32KB） 統合TV
	13:15-14:45			中級			
	15:00-16:30						
	16:45-18:15						
7月24日（金）	10:30-12:00		スクリプト言語	中級	シェルスクリプト	服部 恵美（アメリエフ）	講義資料（PDF:1.8MB） 統合TV
	13:15-14:45						
	15:00-16:30						

②

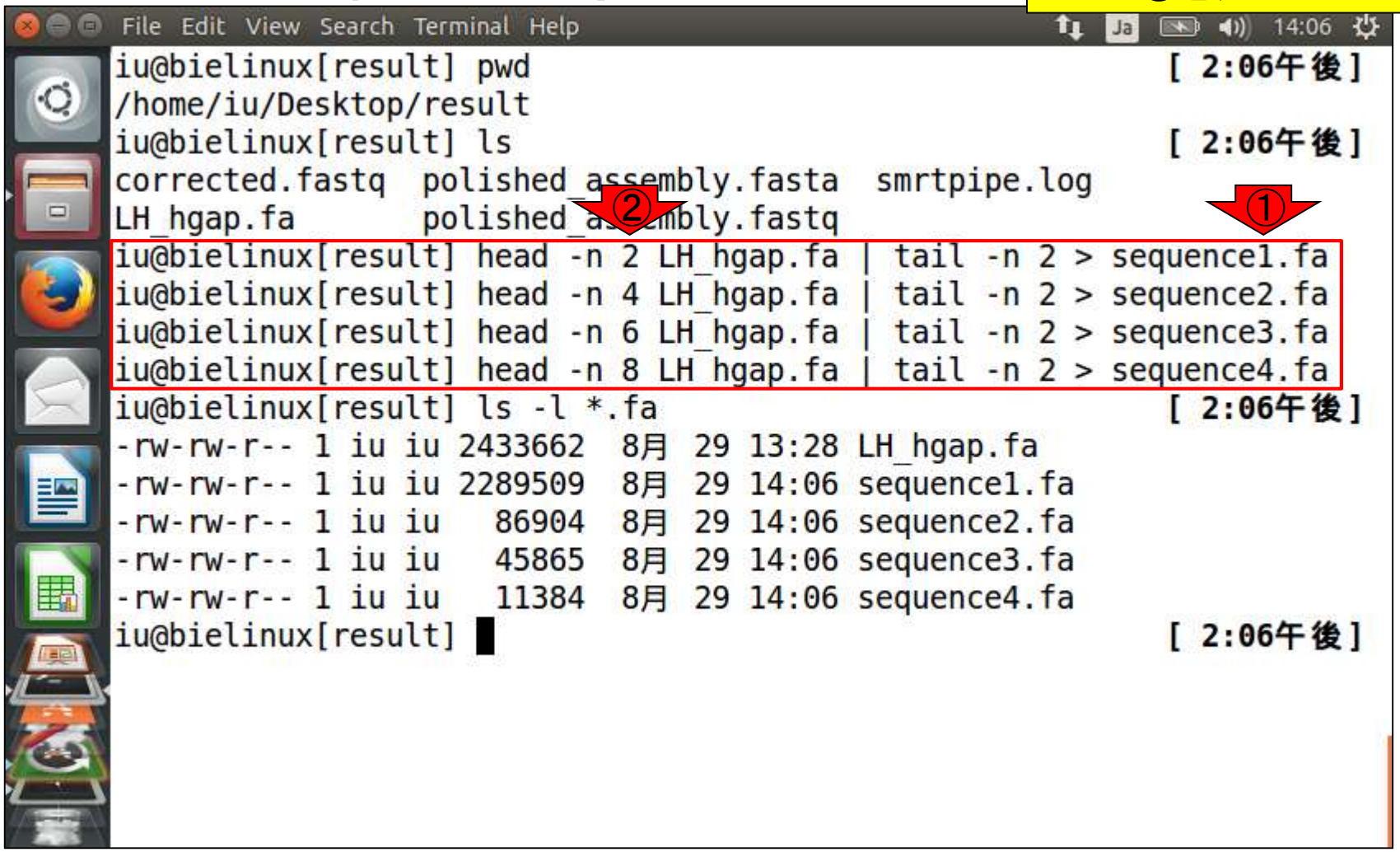
W3-3：戦略を練る

まず、①の部分は2で固定。理由は、出力ファイルがsingle-FASTA形式だから



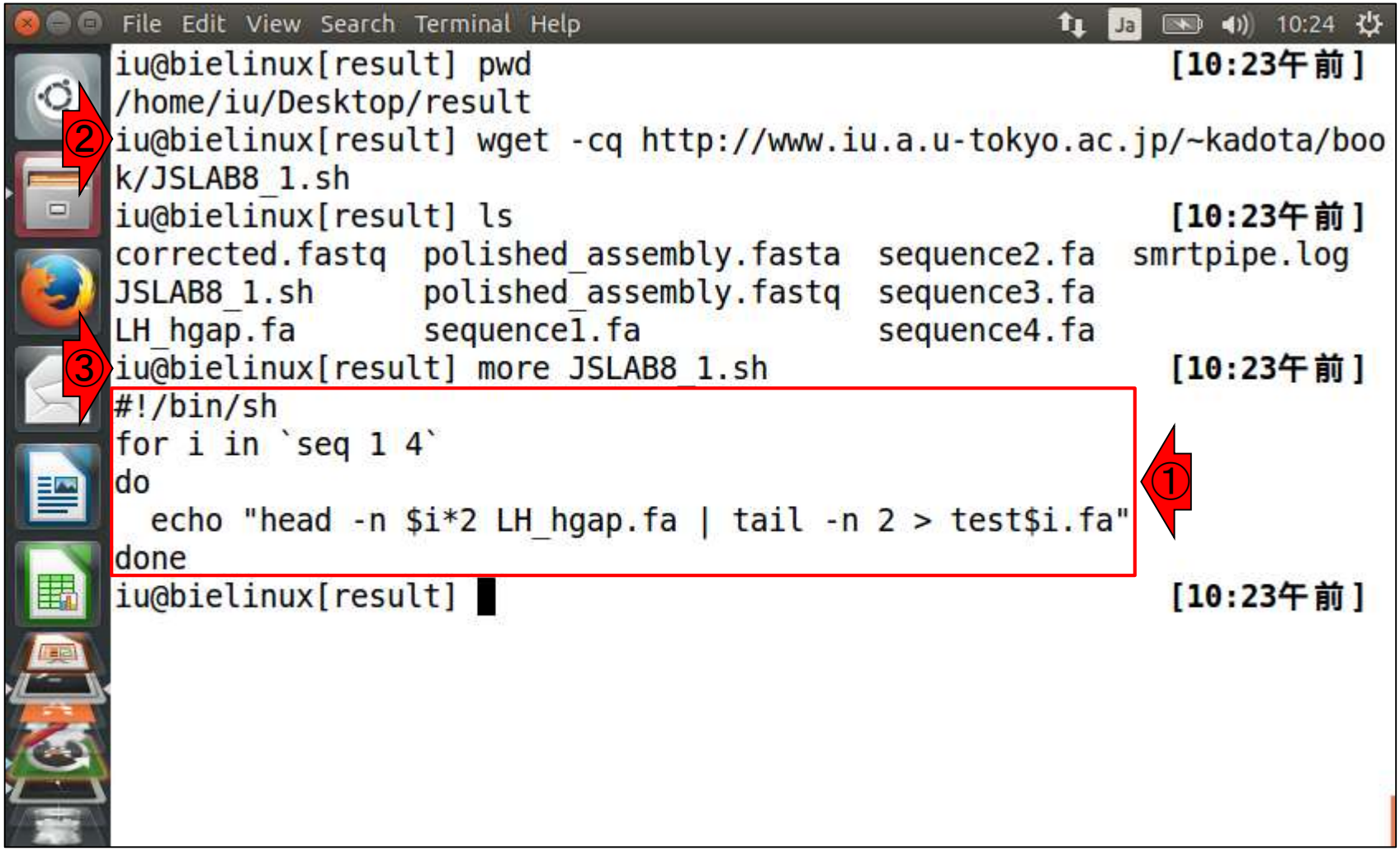
W3-3：戦略を練る

変わっているのは①と②の部分。①1-4までのループを回して、その2倍の値として②を表現できるだろうと妄想する



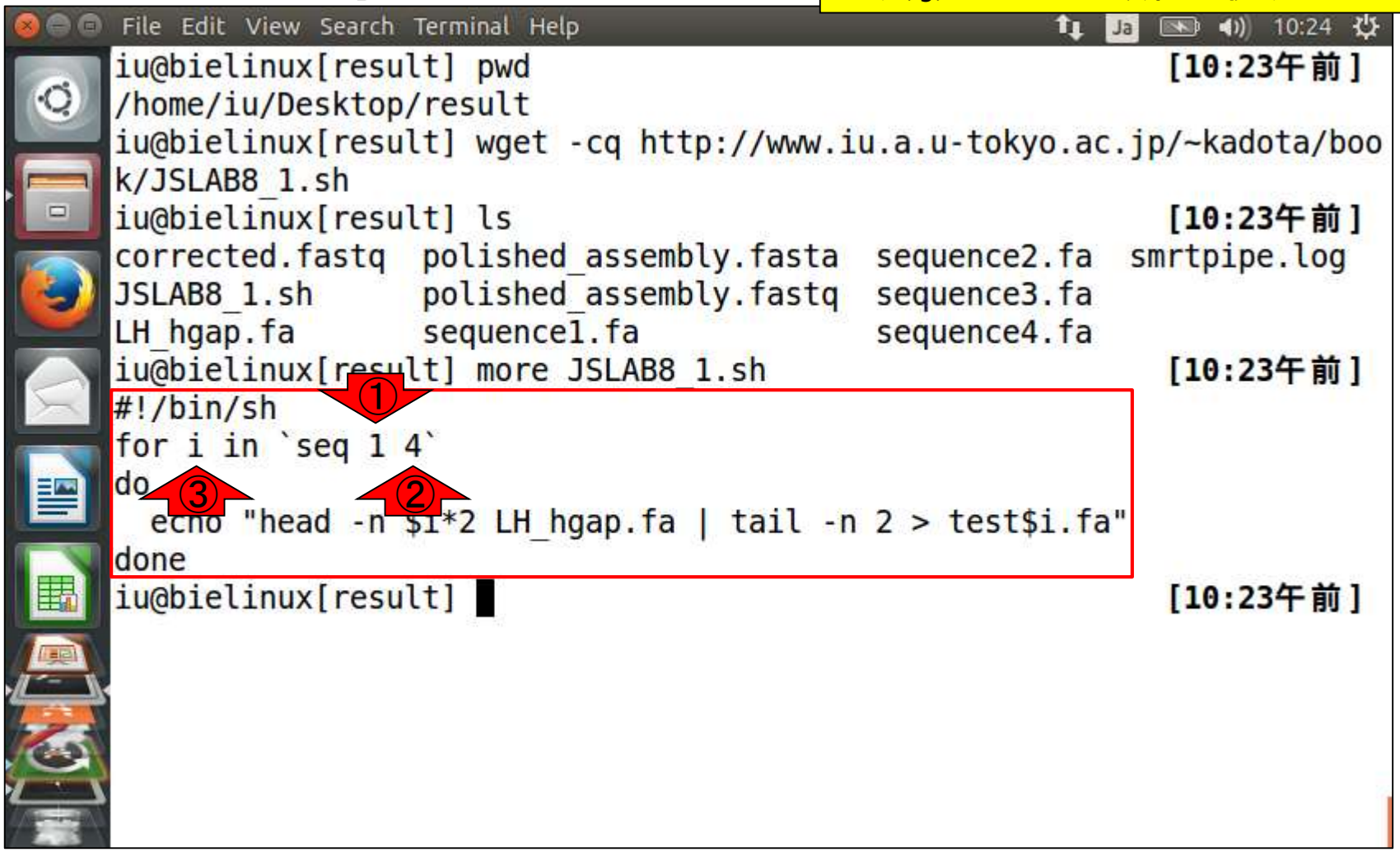
W3-4：基本形

①シェルスクリプトの基本形ファイルJSLAB8_1.shを、②wgetし、③moreで確認



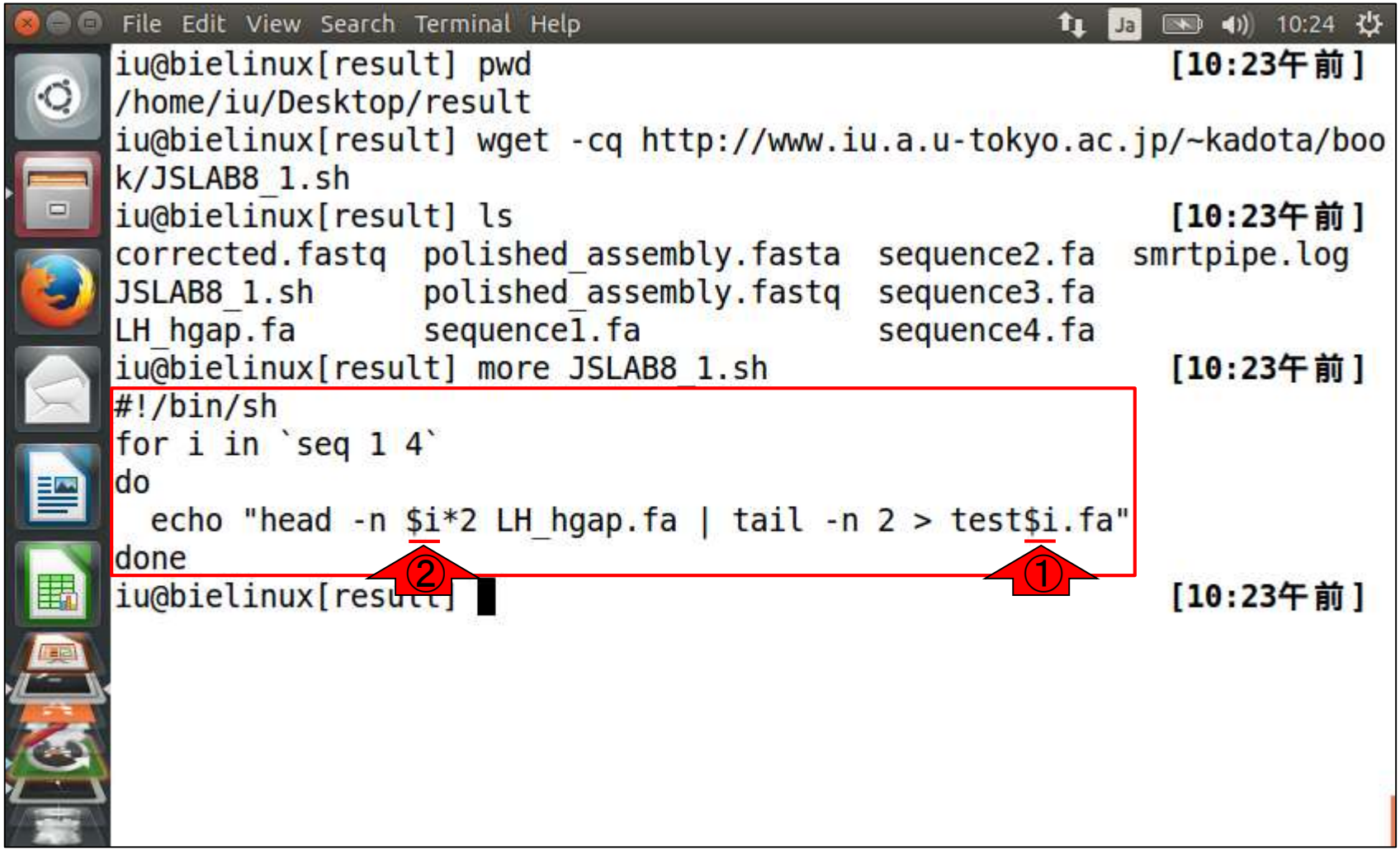
W3-5：解説

①1から②4までの1刻みのループを回し、③iという変数で取り扱う。別にiでなくてもよいが、変数名として、i, j, kといった順番で使うヒトも一定数存在する



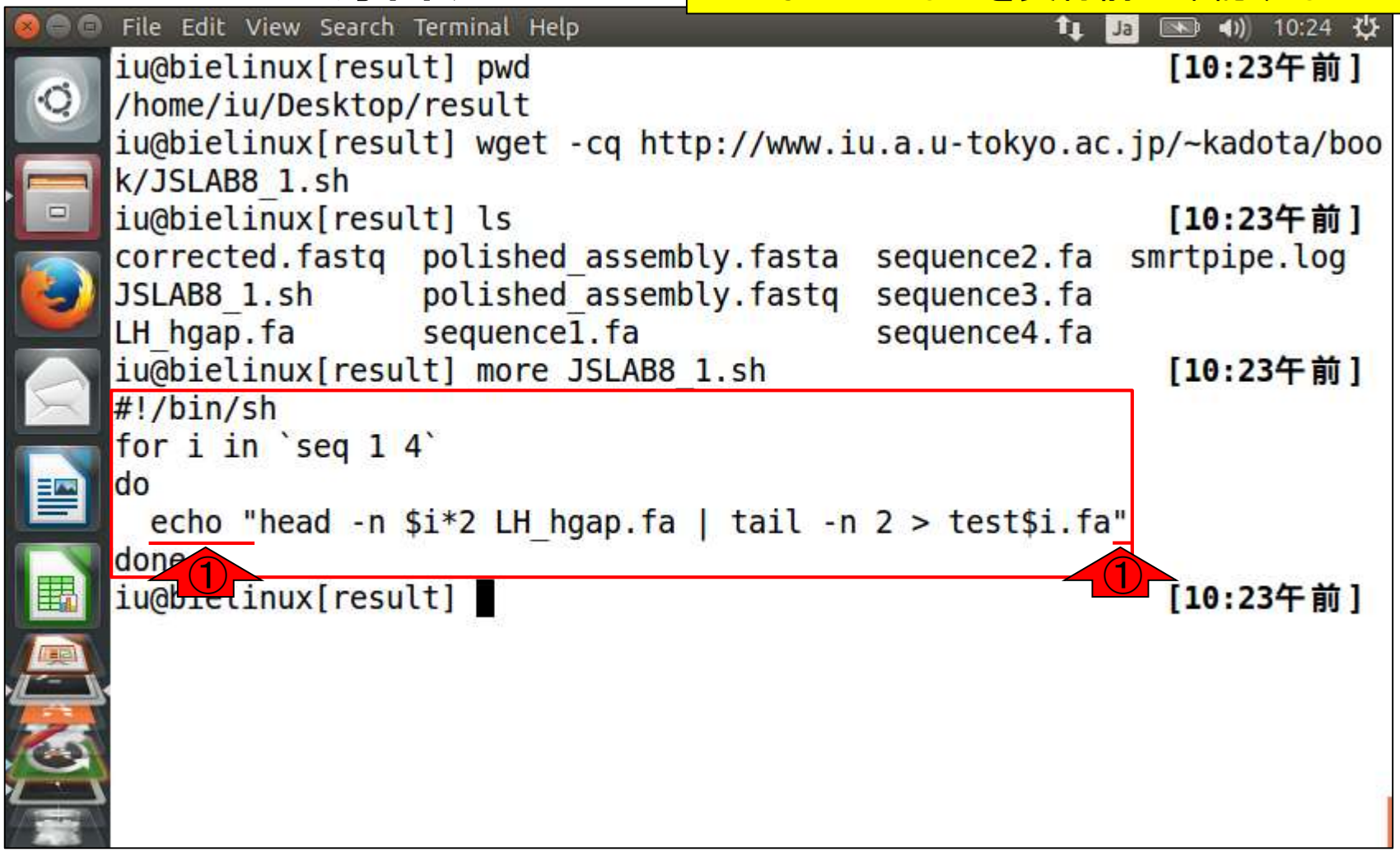
W3-5：解説

変数iは、$iとして取り扱う。つまり$を追加する。なぜ?という類のものではなく、お約束です



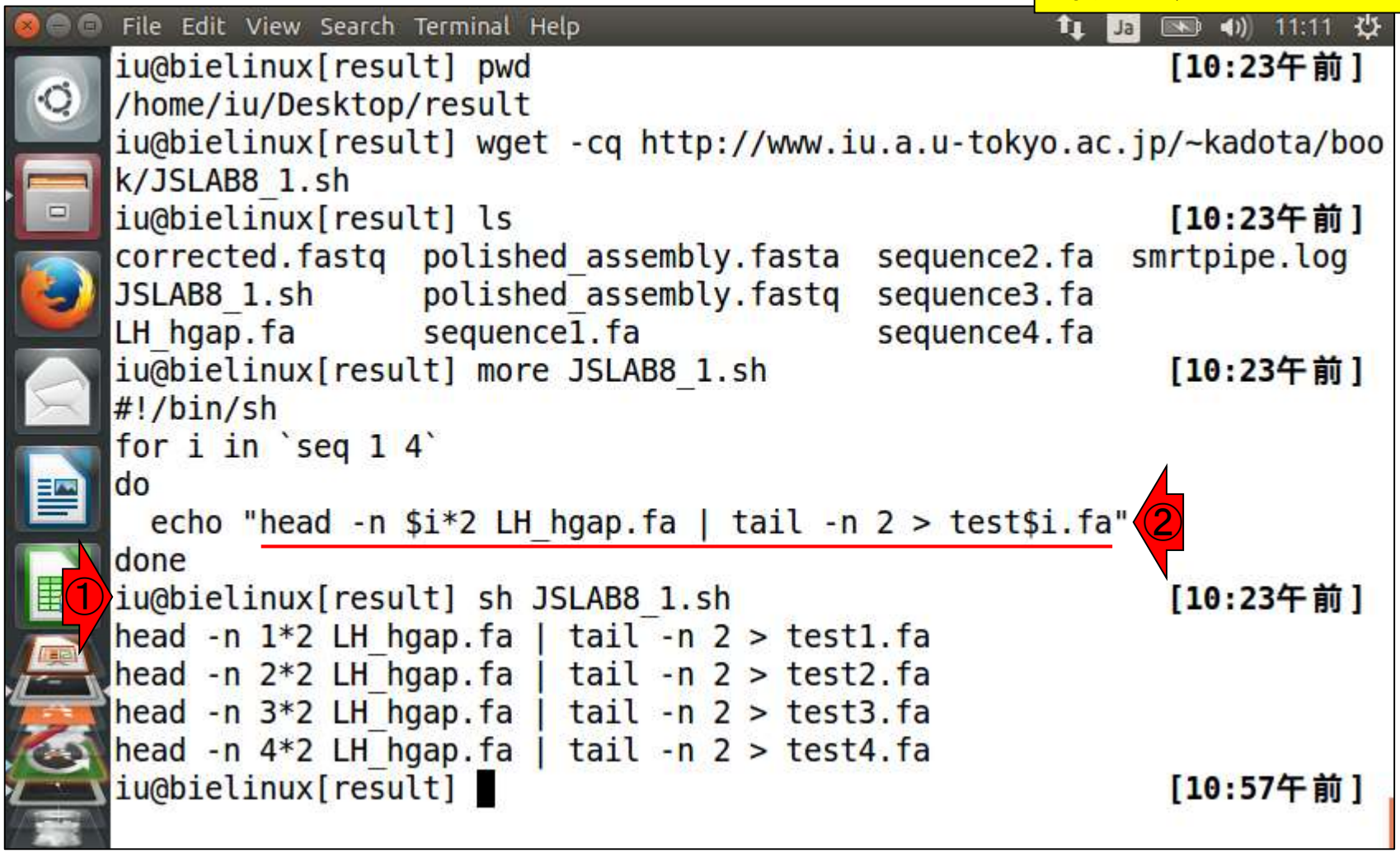
W3-5：解説

①「echo “…”」で囲っていますが、これはLinuxコマンドのechoであり、“”内部の実際に実行したいコマンドがどんな感じになっているかを実行前に確認するためにつけています



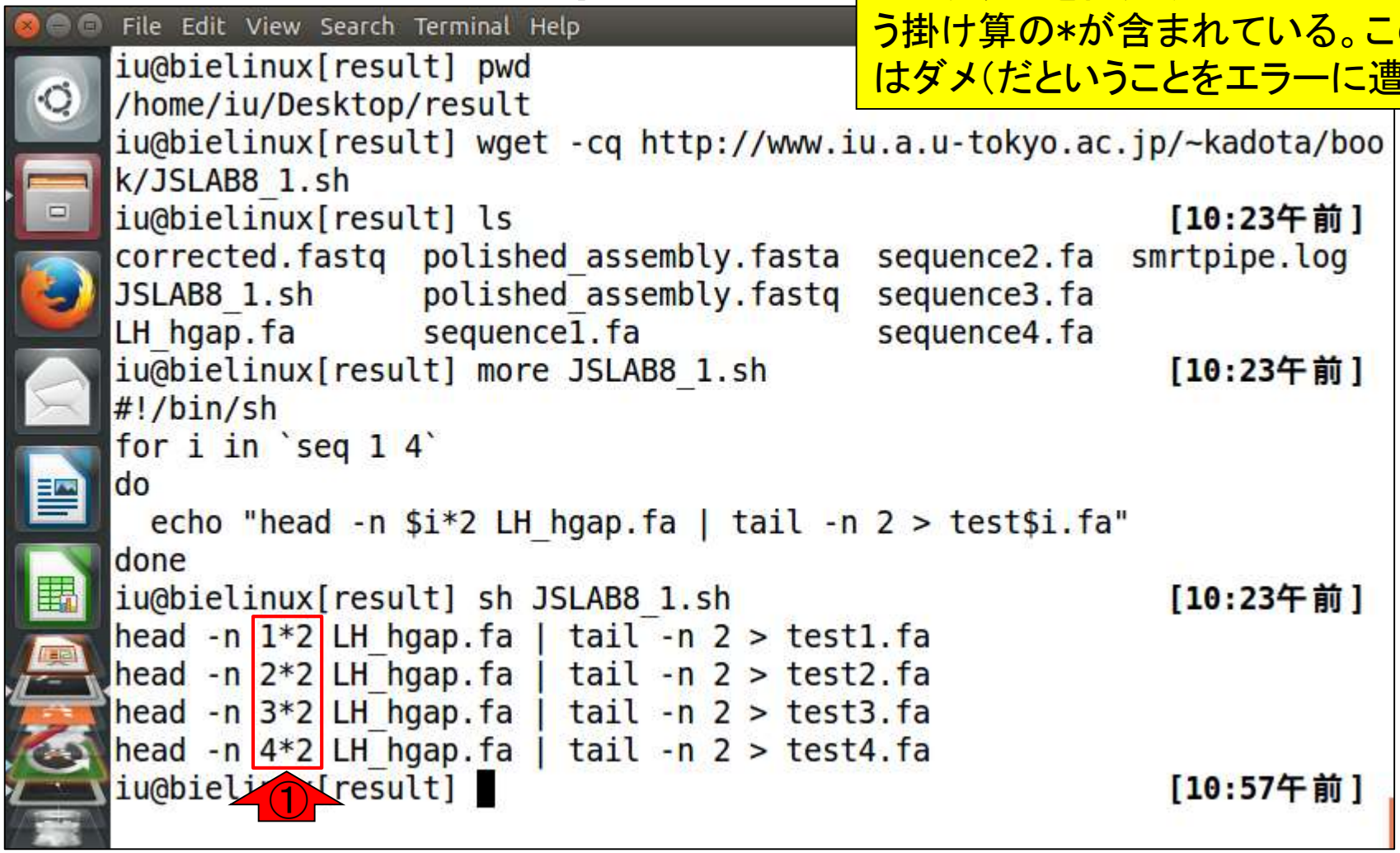
W3-6：echoで確認

①shコマンドでJSLAB8_1.shを実行。②echoで囲った中身に相当する、赤下線部分が表示されていることがわかる



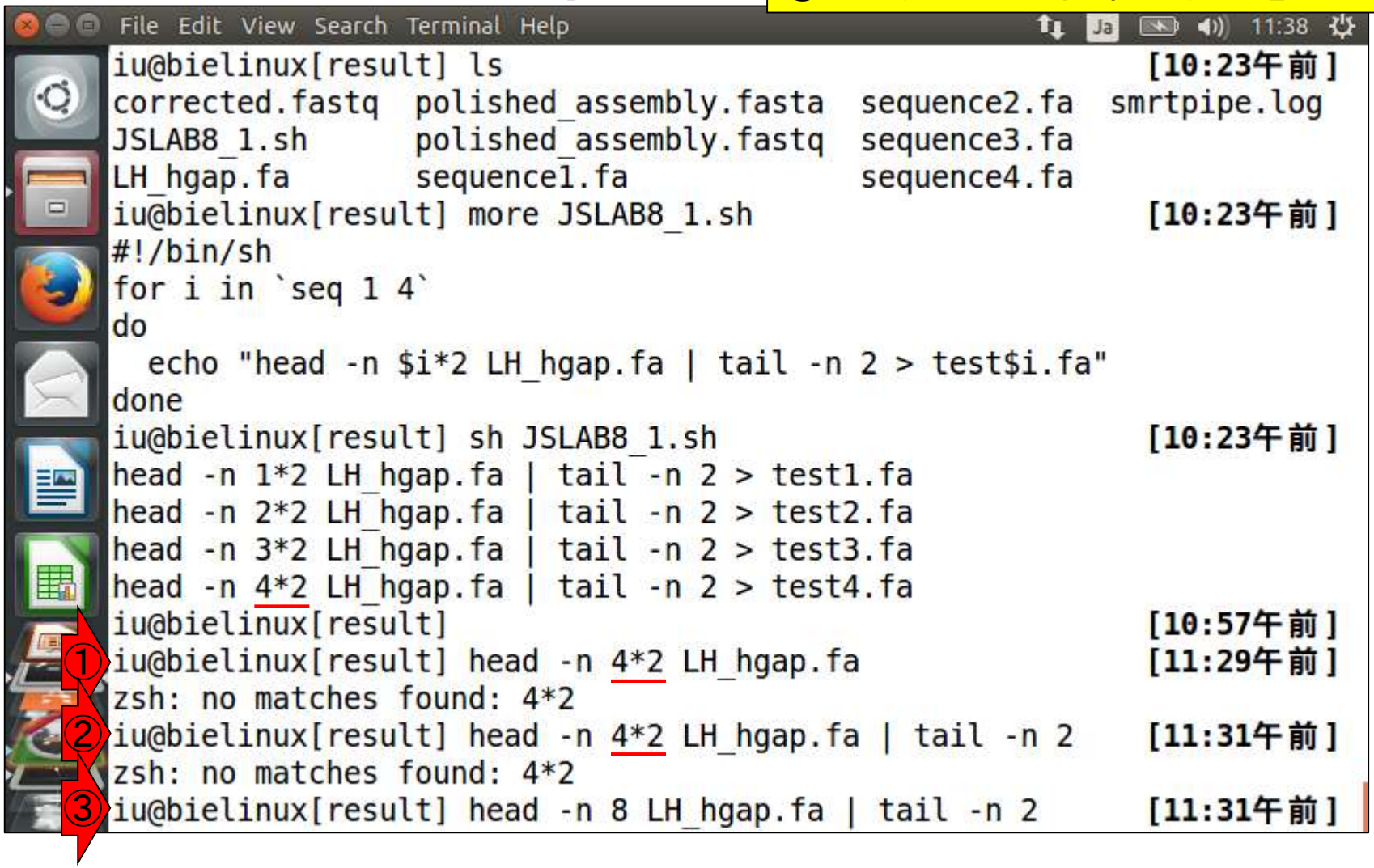
W3-7：head部分

実際に全体を実行してエラーに遭遇してもよいが一応説明。headコマンド部分の①では最初のy行という数値を指定するところだが、1*2とか4*2という掛け算の*が含まれている。このような指定法はダメ(だということをエラーに遭遇して学習する)



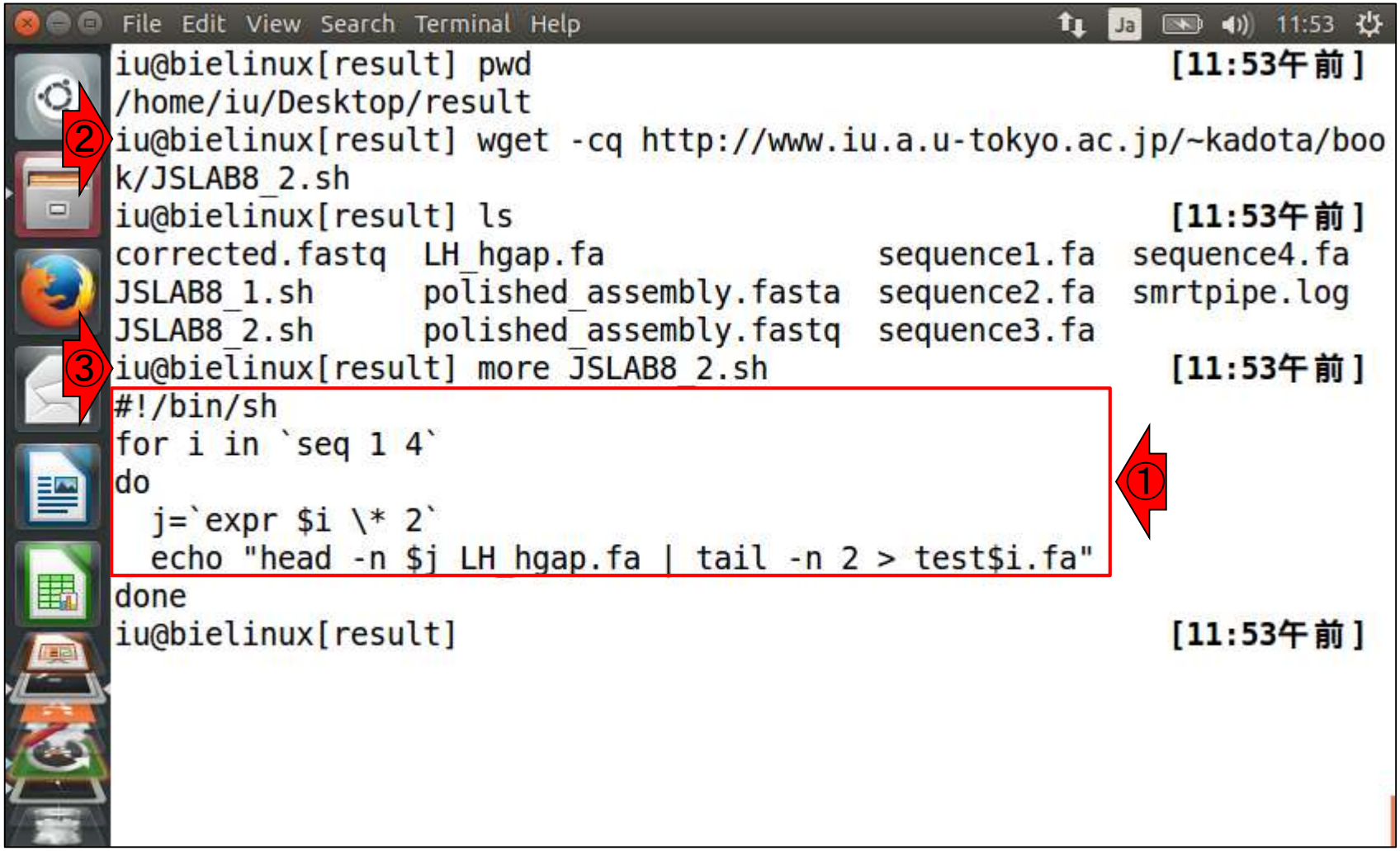
W3-7：head部分

①や②を実際に実行してみても、赤下線部分が原因でエラーが出ます。エラーが出ないようにするには、③のように4*2の結果である8を与える必要があります



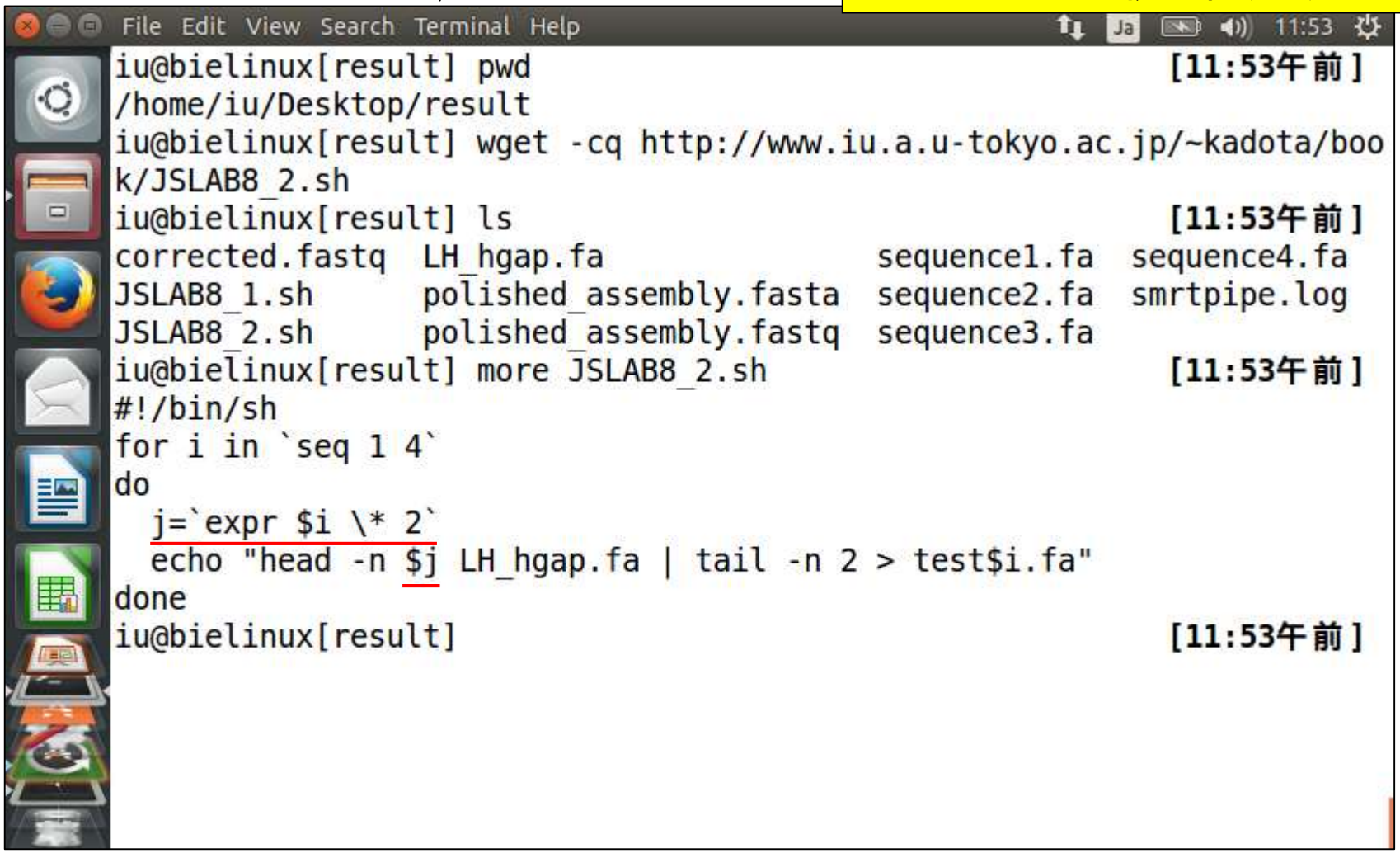
W3-8：発展形

①シェルスクリプトの発展形ファイルJSLAB8_2.shを、②wgetし、③moreで確認



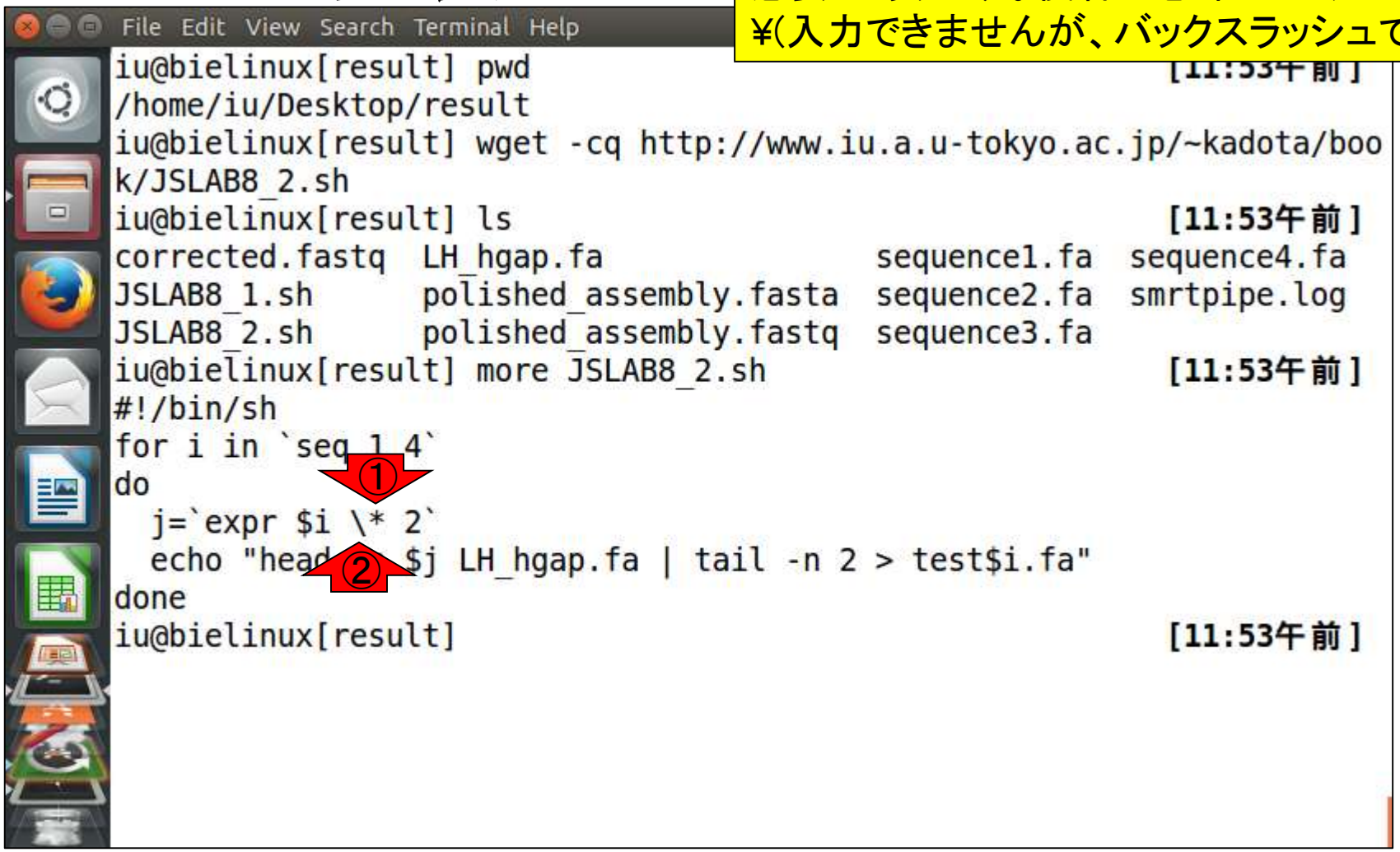
W3-8：発展形

JSLAB8_1.shとの違いは、赤下線部分のみ。jという別の変数を用いて$i*2を表現したいだけだが、シェルスクリプトの掟に従うと、このようになります



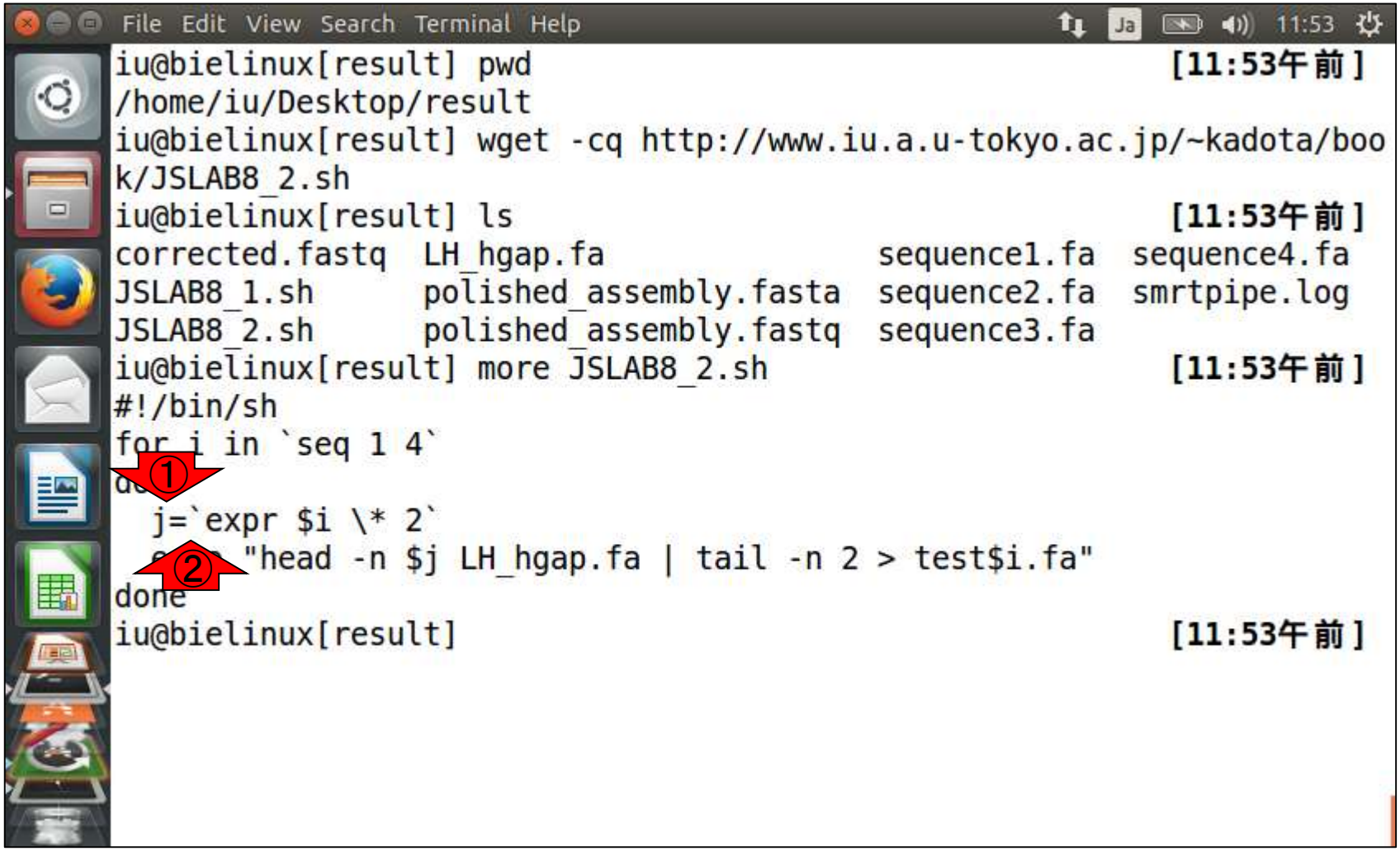
W3-8：発展形

①は$i*2の*に相当する部分ですが、「なんでもよい、というワイルドカードの*」と「掛け算の*」の意味を区別する必要があります。後者の意味として用いたい場合に、②¥(入力できませんが、バックスラッシュです)を添えます



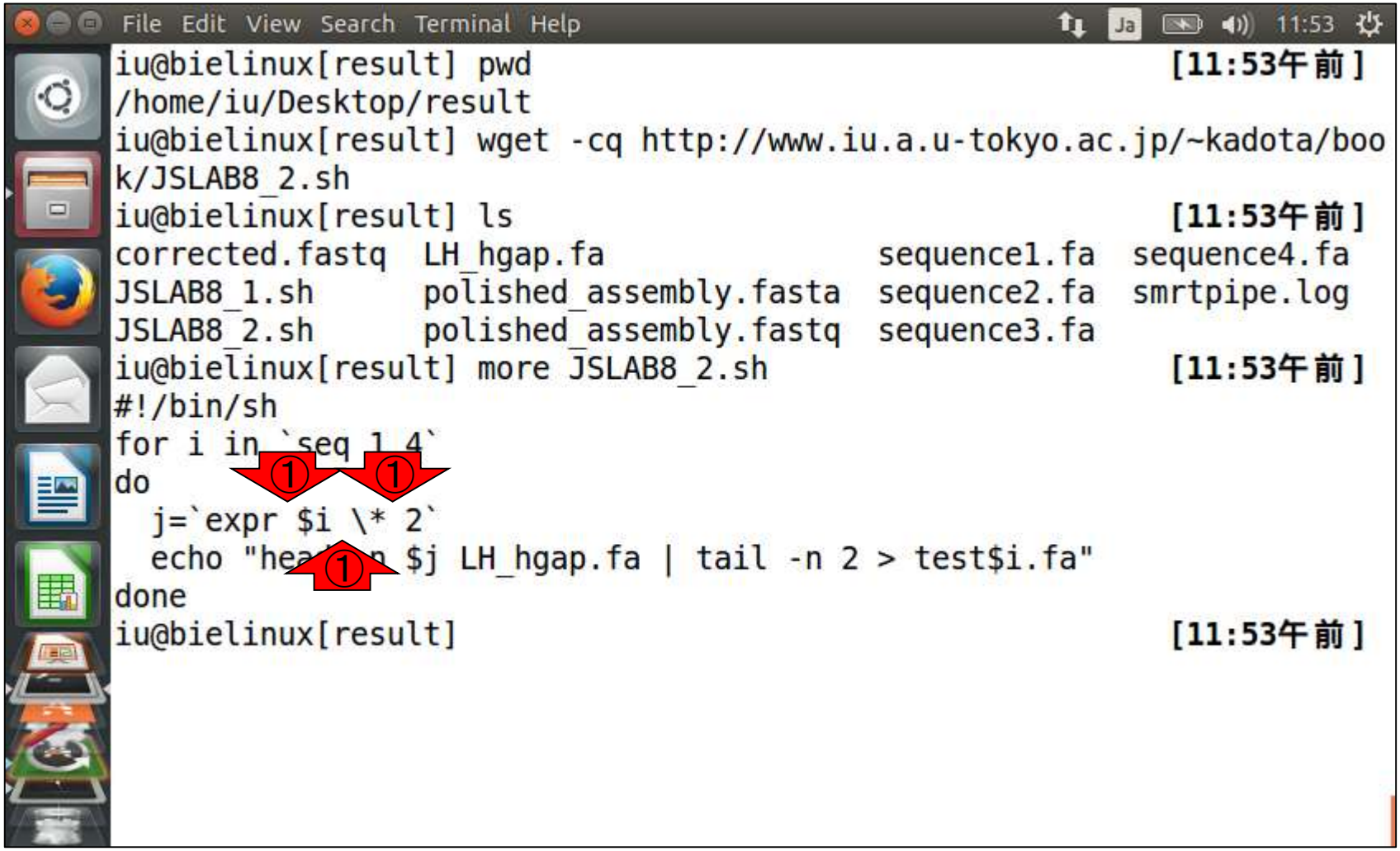
W3-8：発展形

掟の続き。①や②のところにスペースを入れてはいけません



W3-8：発展形

掟の続き。①のところにはスペースを入れないといけません



W3-8：発展形

①の記号は、②Shiftキーを押しながら、③@のキーを押すと出ます

File Edit View Search Terminal Help

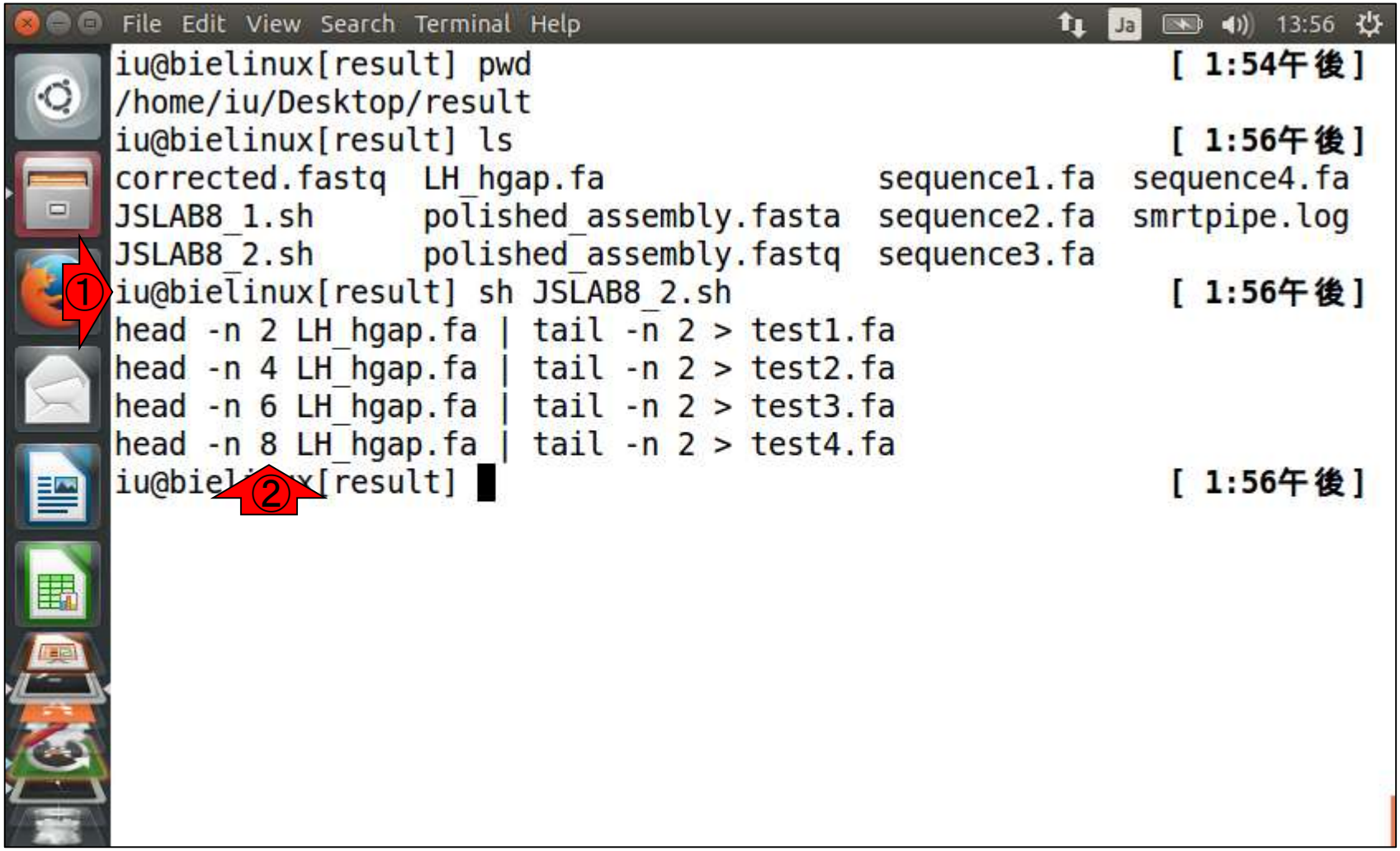
```
iu@bielinux[result] pwd                                    [11:53午前]
/home/iu/Desktop/result
iu@bielinux[result] wget -cq http://www.iu.a.u-tokyo.ac.jp/~kadota/boo
k/JSLAB8_2.sh
iu@bielinux[result] ls                                     [11:53午前]
corrected.fastq  LH_hgap.fa                  sequence1.fa  sequence4.fa
JSLAB8_1.sh      polished_assembly.fa
JSLAB8_2.sh      polished_assembly.fa
iu@bielinux[result] more JSLAB8_2.sh
#!/bin/sh
for i in `seq 1 4`
do
  j=`expr $i \* 2`
  echo "head -n $j LH_hgap.fa | tail
done
iu@bielinux[result]
```

①

①

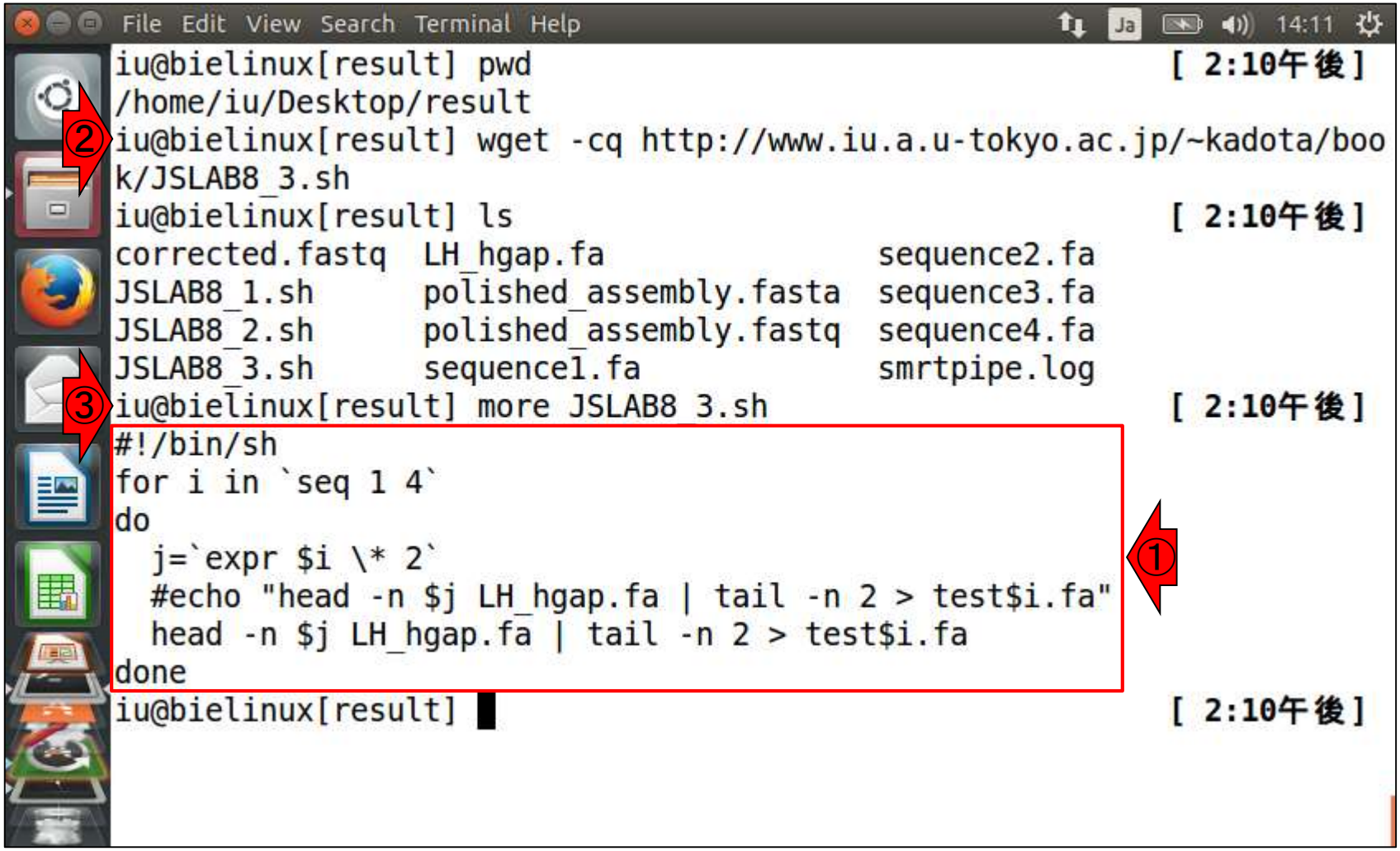
W3-9：echoで確認

①shコマンドでJSLAB8_2.shを実行。
②意図通りになっていることがわかる



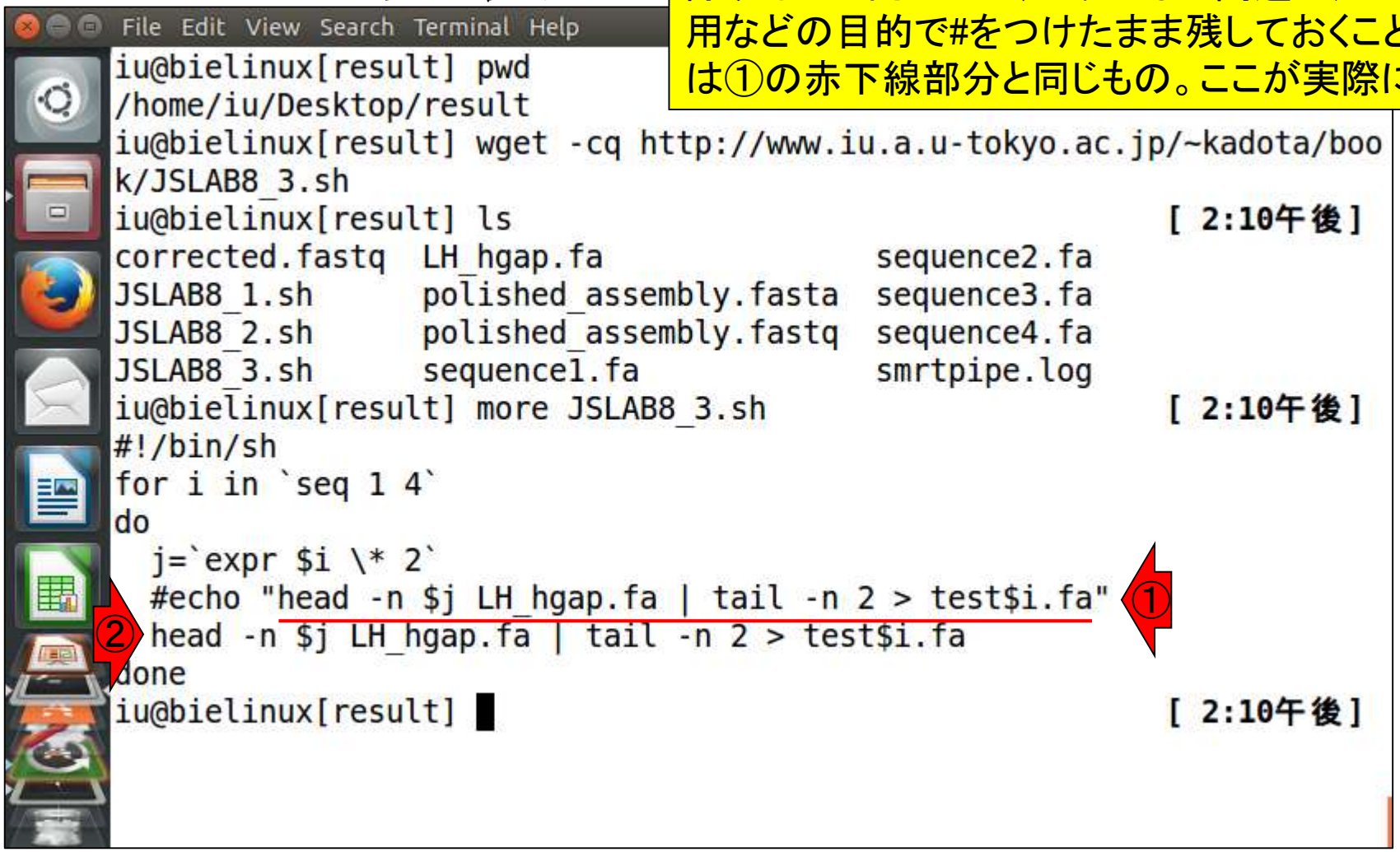
W3-10：発展形2

①シェルスクリプトの発展形2ファイル
JSLAB8_3.shを、②wgetし、③moreで確認



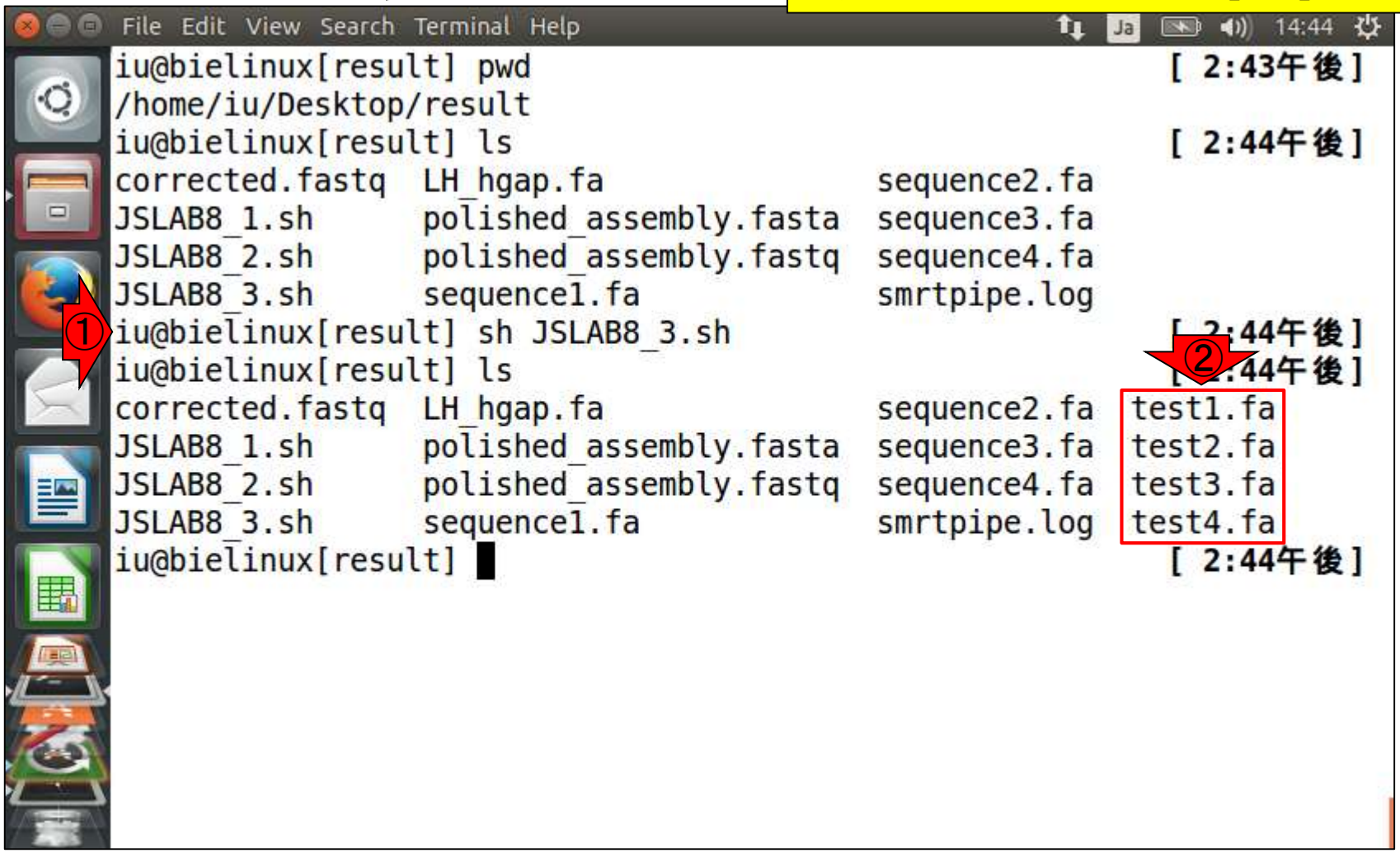
W3-10：発展形2

JSLAB8_2.shとの違いは、①と②の行の部分。①のecho行頭に#を入れてコメントアウト（実行されないように）している。削除するのと同じだが、エラーなど問題が起こったときの対処用などの目的で#をつけたまま残しておくことはよくやる。②は①の赤下線部分と同じもの。ここが実際に実行される部分



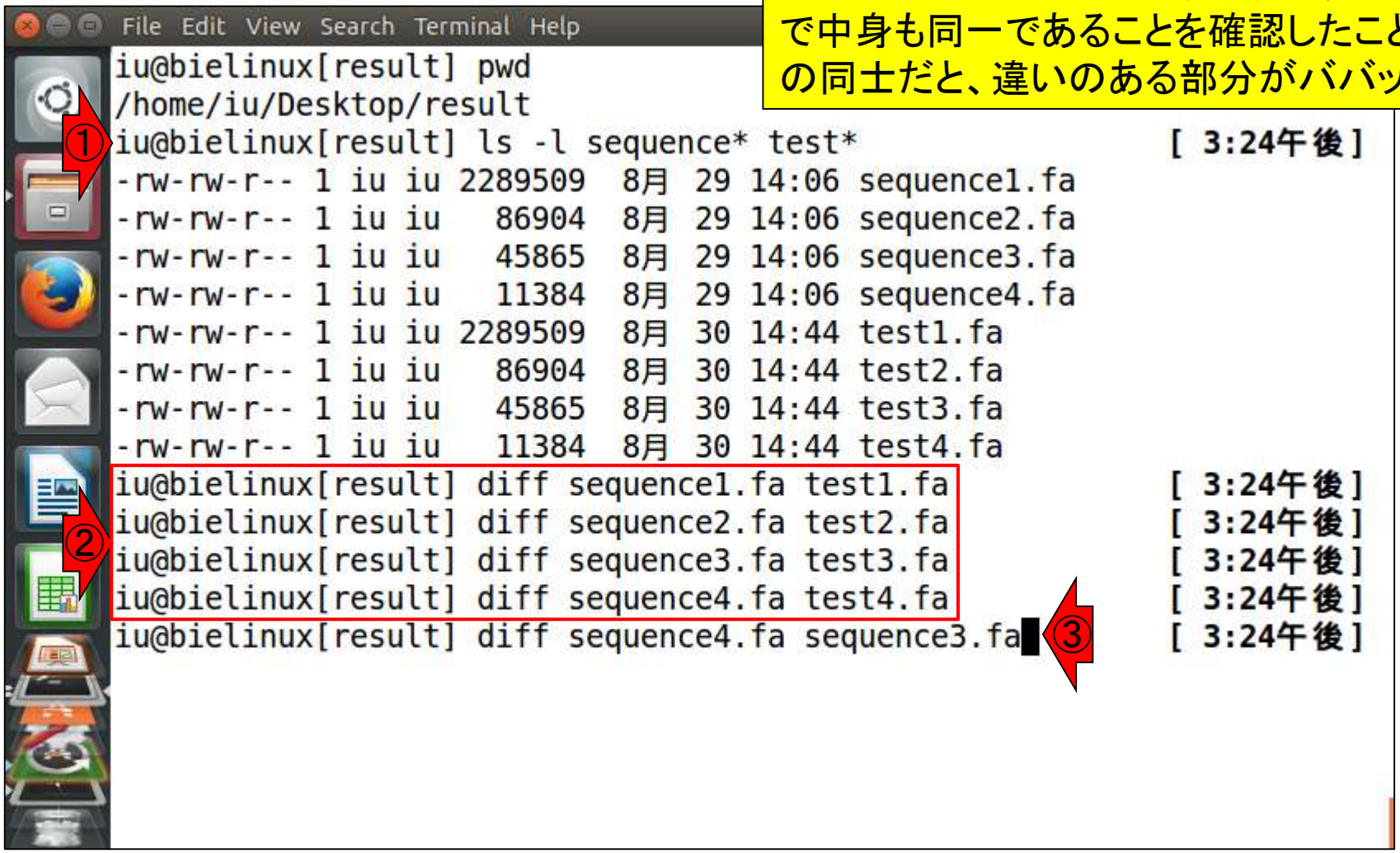
W3-11：実行

①「sh JSLAB8_3.sh」を実行。echoをコメントアウトしているので、意図通り何も表示されない。また、②イメージ通りの出力ファイルtest[0-9].faが作成されている



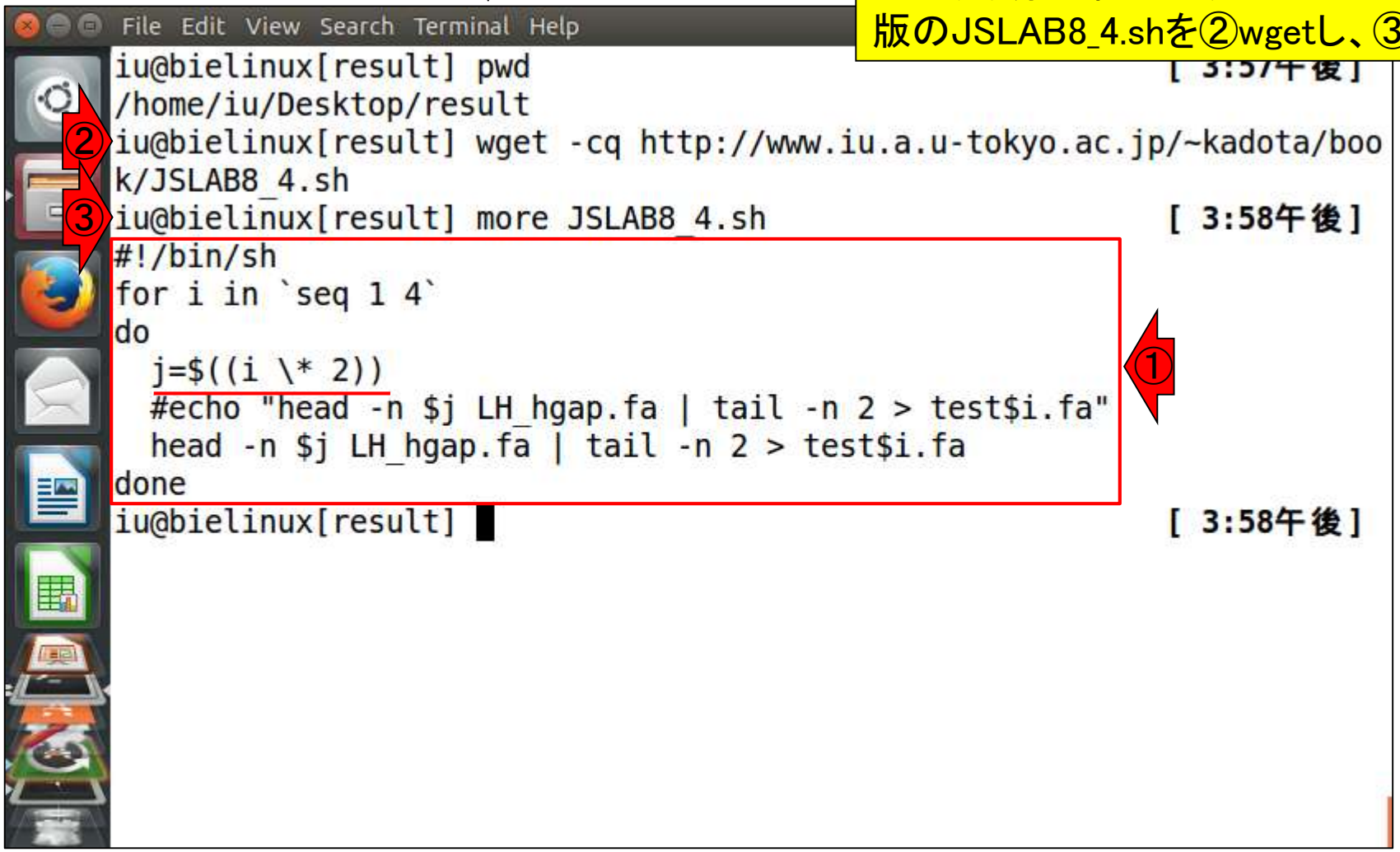
W3-12：確認

①「ls -l」で詳細情報を表示。ファイルサイズの観点からは、sequence[0-9].faとtest[0-9].faは全く同じ。②念のためdiffコマンドでも確認。何も表示されていないので中身も同一であることを確認したことになる。③違うもの同士だと、違いのある部分がババッと表示されます



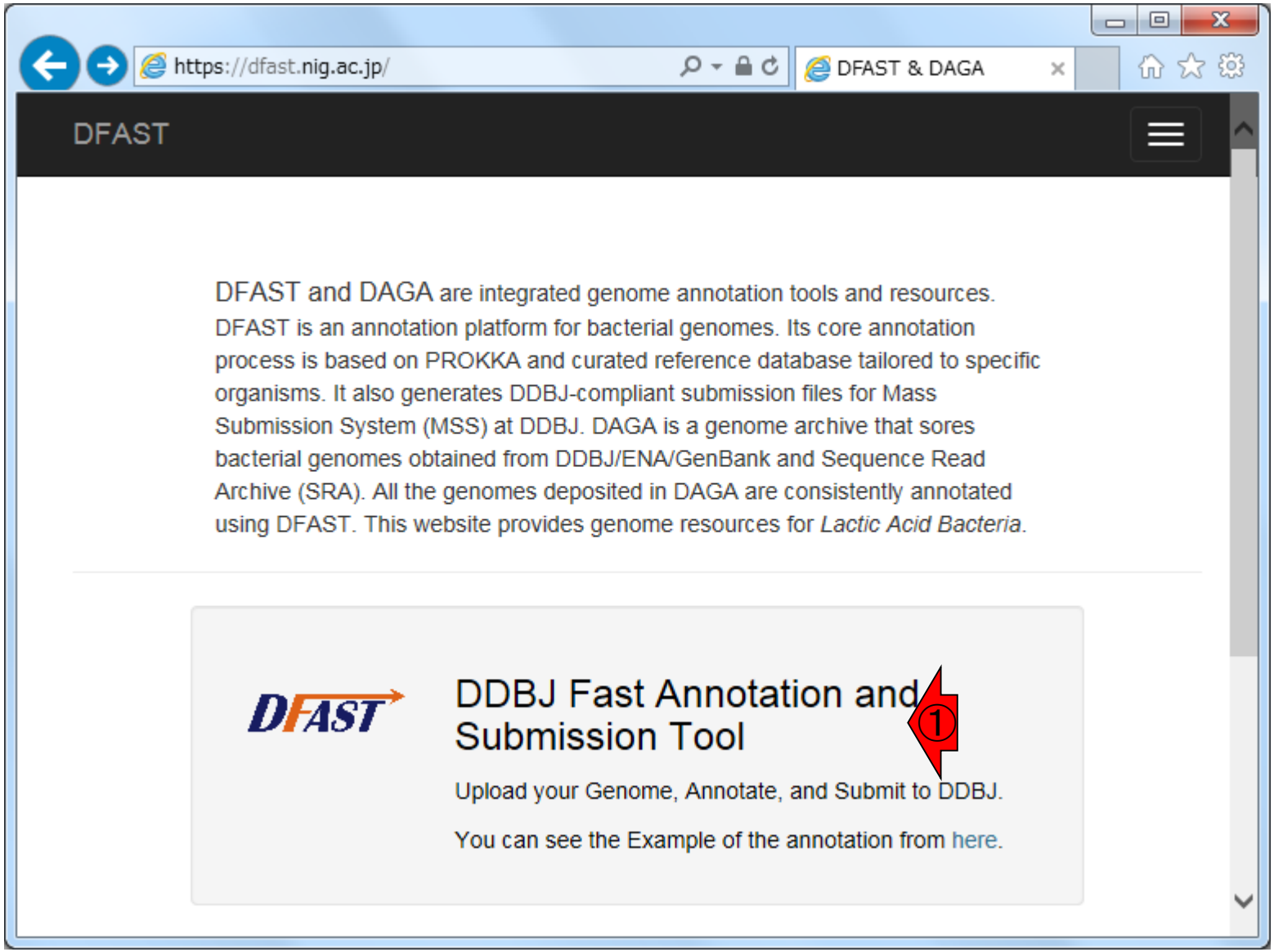
W3-13：発展形3

「expr シェル 遅い」などでググればわかりますが、JSLAB8_3.sh内で使われているexprは遅いらしいです。赤下線のように書くといいらしい。①高速版のJSLAB8_4.shを②wgetし、③more。後は省略



DFASTのトップ画面。①をクリック

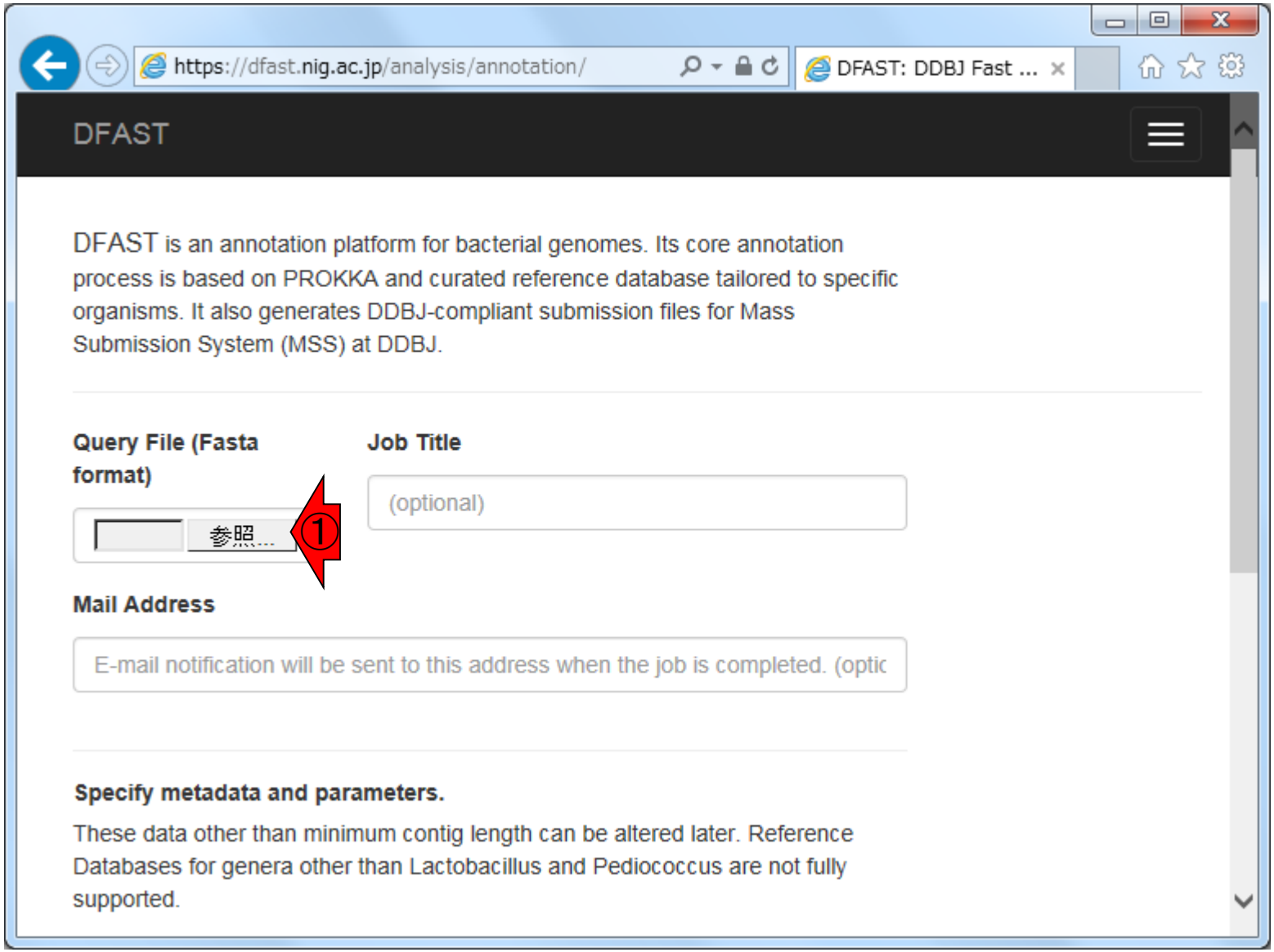
W4-1：DFAST



Tanizawa et al., *Biosci Microbiota Food Health*, **35**: 173-184, 2016

W4-1：DFAST

①のところで、参照ボタンを押して、アノテーションしたいmulti-FASTAファイル(LH_hgap.fa)を指定



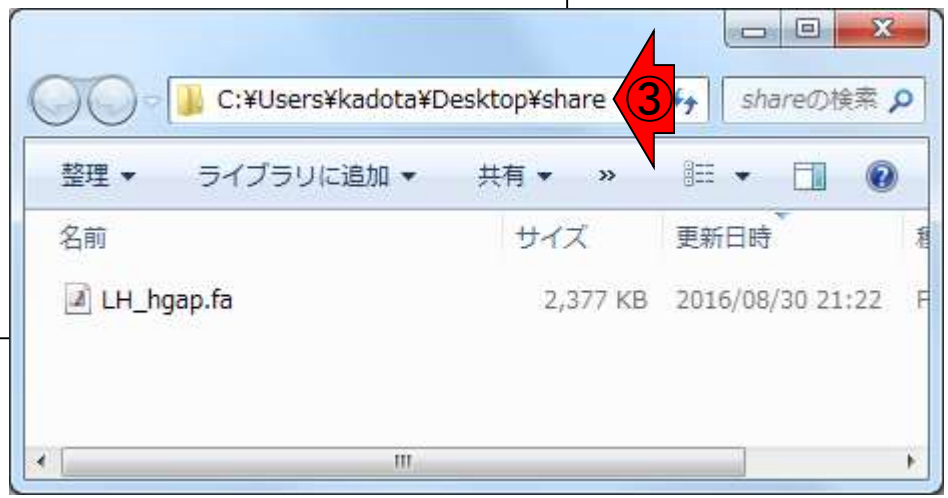
W4-2：LH_hgap.fa

①アノテーションしたいmulti-FASTAファイル(LH_hgap.fa)を共有フォルダにコピー。②共有フォルダ内の状況はヒトそれぞれだが、最低限LH_hgap.faが見えていればよい。(私のWindows環境では)ホストOS上の絶対パスとして③のところにLH_hgap.faが見えるので、これをアップロードします

```
iu@bielinux[~/Desktop/result]
iu@bielinux[result] pwd
/home/iu/Desktop/result
iu@bielinux[result] ls
corrected.fastq  LH_hgap.fa                 sequence3.fa  test3.fa
JSLAB8_1.sh      polished_assembly.fasta    sequence4.fa  test4.fa
JSLAB8_2.sh      polished_assembly.fastq    smrtpipe.log
JSLAB8_3.sh      sequence1.fa               test1.fa
JSLAB8_4.sh      sequence2.fa               test2.fa
iu@bielinux[result] cp LH_hgap.fa ../mac_share          [ 9:21午後]
iu@bielinux[result] ls -l ../mac_share                  [ 9:22午後]
total 2377
-rwxrwxrwx 1 iu iu 2433662  8月 30 21:22 LH_hgap.fa
iu@bielinux[result]                                     [ 9:22午後]
```

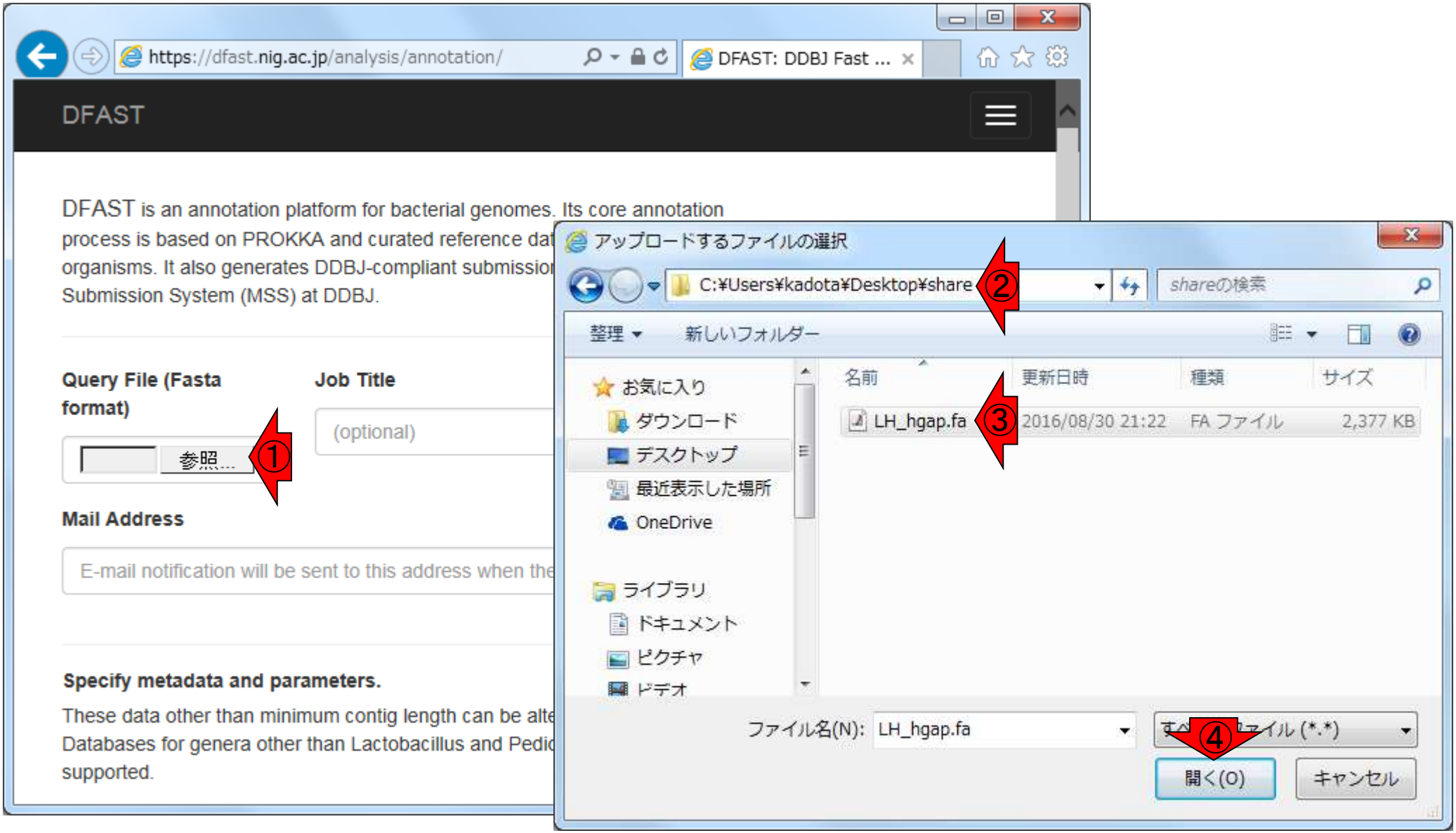
①

②



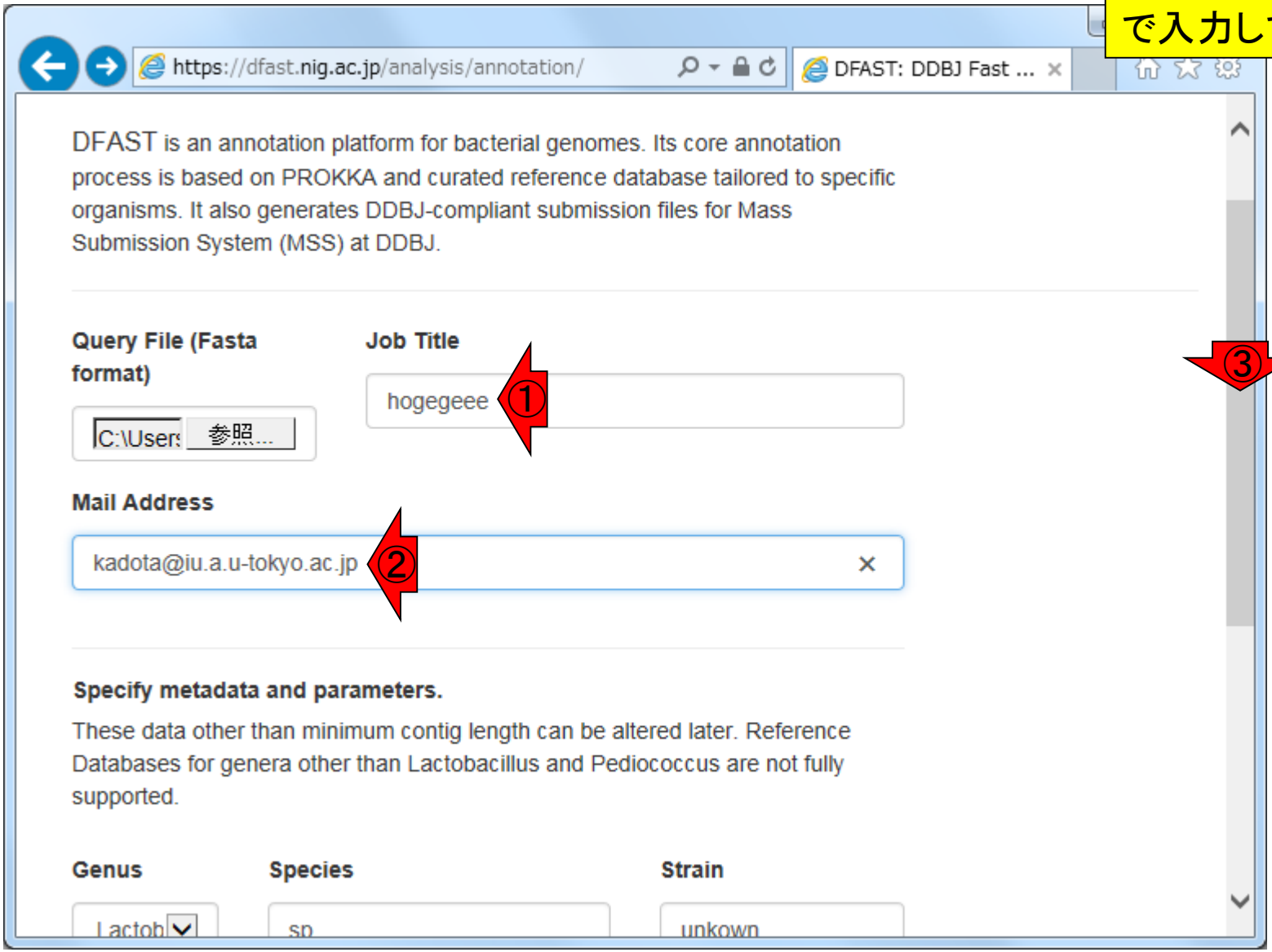
W4-3：アップロードと実行

①参照ボタン、②に移動して、③目的のファイル(LH_hgap.fa)を指定して、④開く



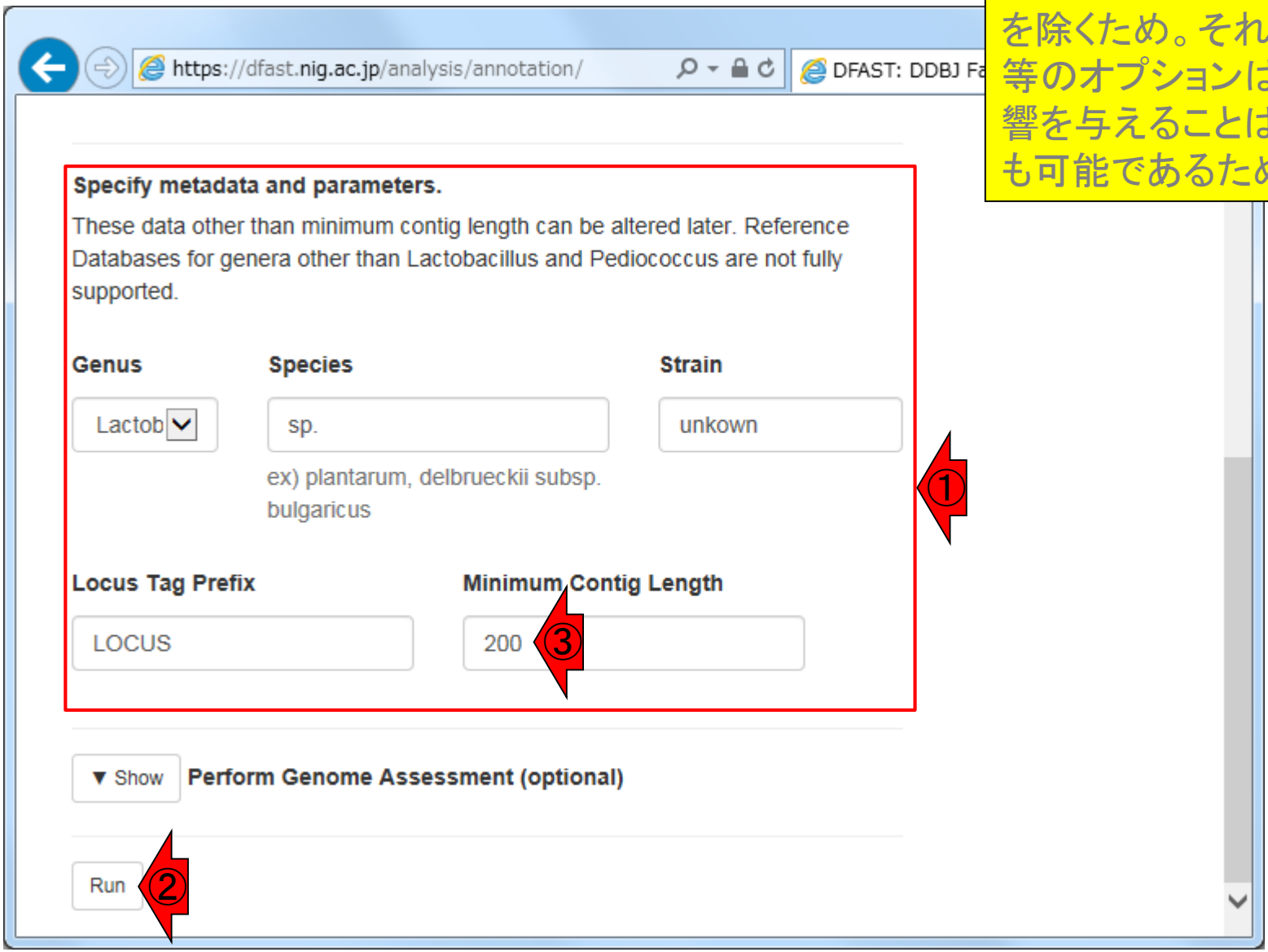
W4-3：アップロードと実行

①ジョブタイトル(hogegeee)をテキトーにつけて、②計算が終わったらメールでお知らせしてくれるので入力して、③ページ下部に移動



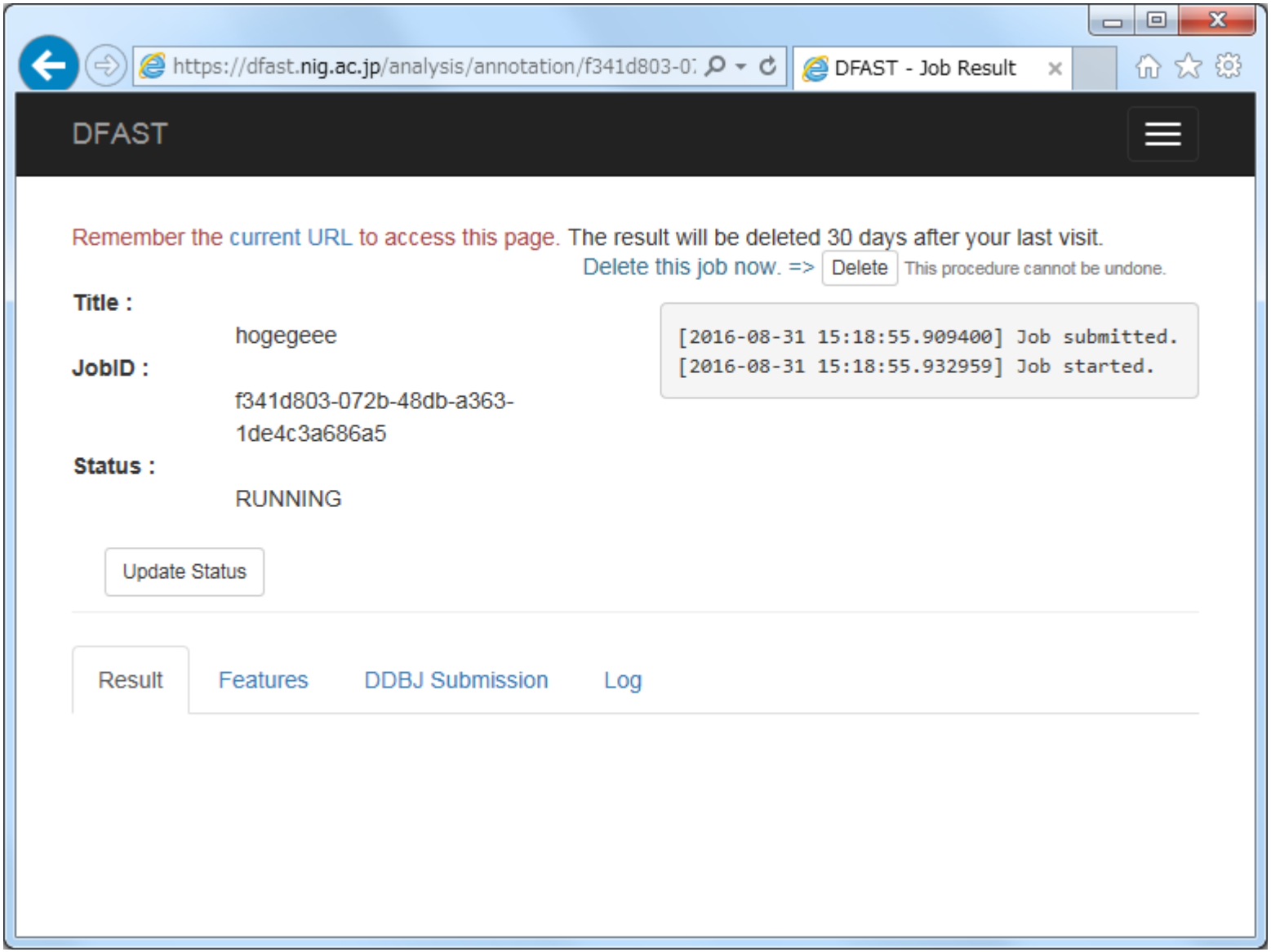
W4-3：アップロードと実行

①いろいろオプション指定できるがとりあえずここは無視して、②Run。③は設定値(デフォルトは200 bp)以下の短い配列を除くため。それ以外の赤枠の属・種名等のオプションはアノテーション結果に影響を与えることはなく、後で変更することも可能であるため、規定値のままでOK



W4-3：アップロードと実行

計算が始まったようだ。とりあえず計算終了メールが来るまで思考停止



W4-4：計算終了

計算終了メールが届く。このときは、確かに5分以内に計算が終了。①にアクセス

2016/08/31 (水) 15:22

DFAST Report <dfast@nig.ac.jp>

[DFAST] Your requested job completed.

宛先 kadota@iu.a.u-tokyo.ac.jp

Your requested job has completed.

Job Title : hogegeee

Job ID : f341d803-072b-48db-a363-1de4c3a686a5

Submitted at :2016-08-31 15:18:55.909400.

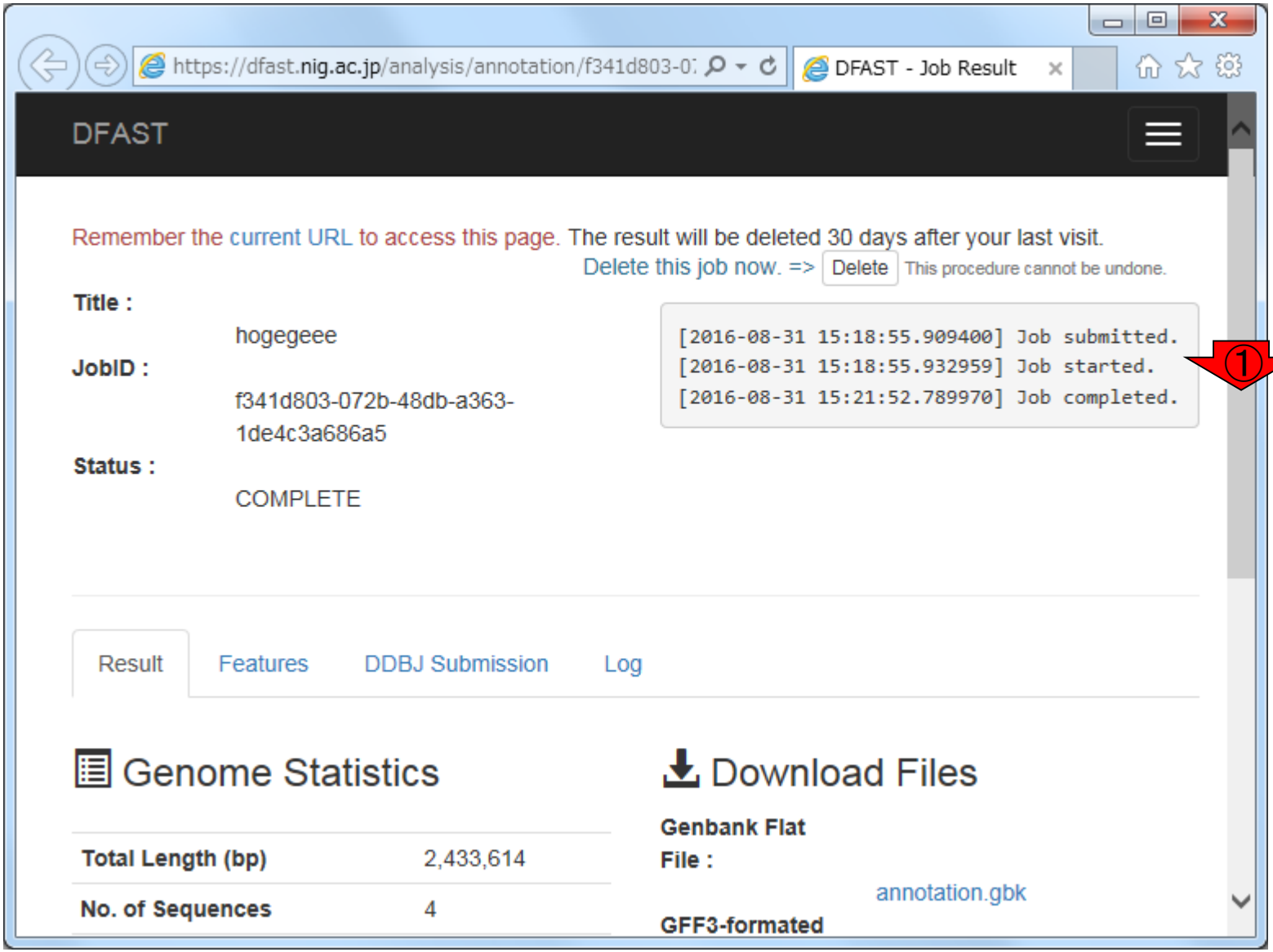
Please visit the link below to check the result.

https://dfast.nig.ac.jp/analysis/annotation/f341d803-072b-48db-a363-1de4c3a686a5 ①

DFAST <dfast@nig.ac.jp>

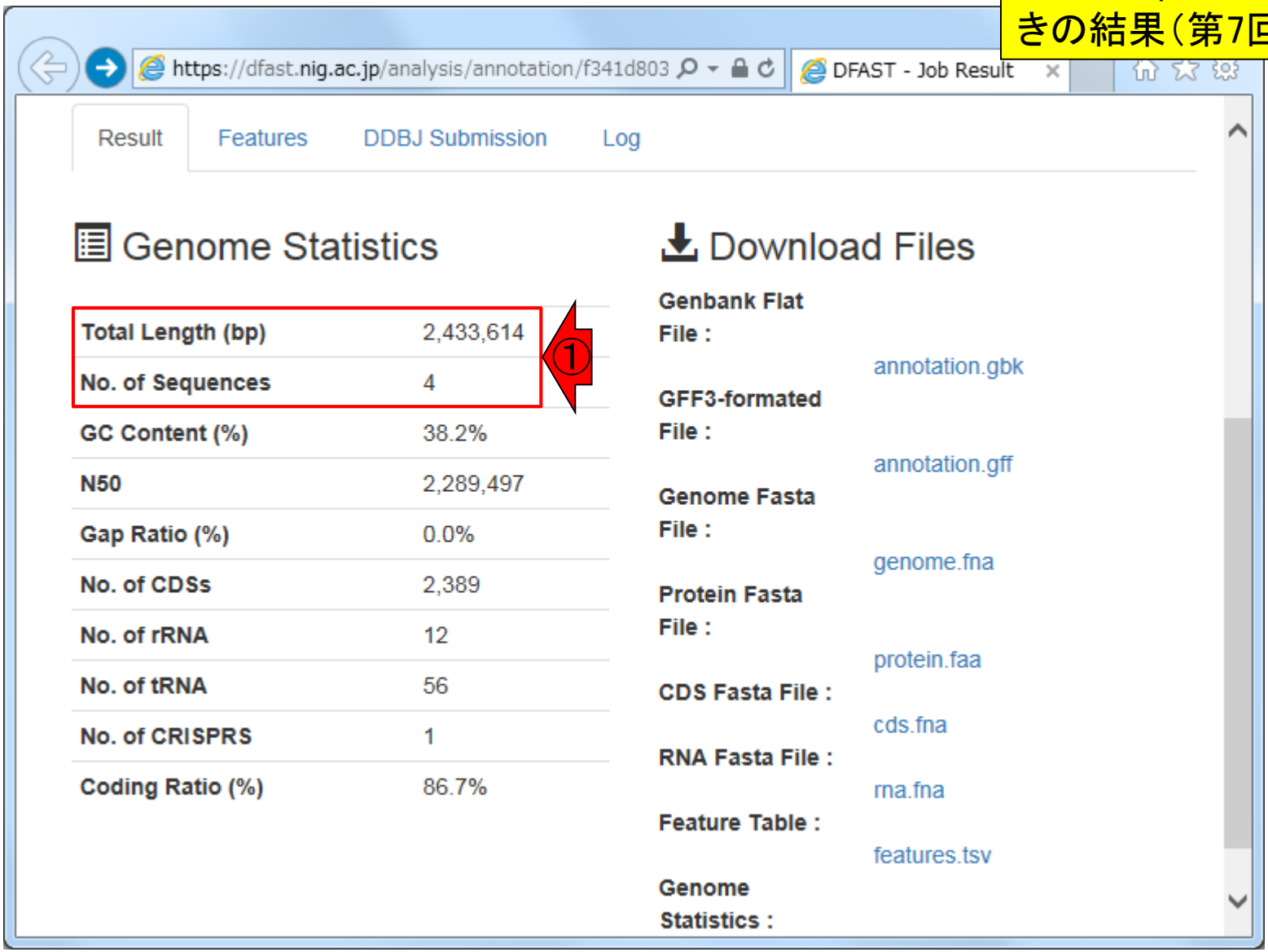
W4-5：結果を眺める

計算結果が見えている。①ページ下部に移動



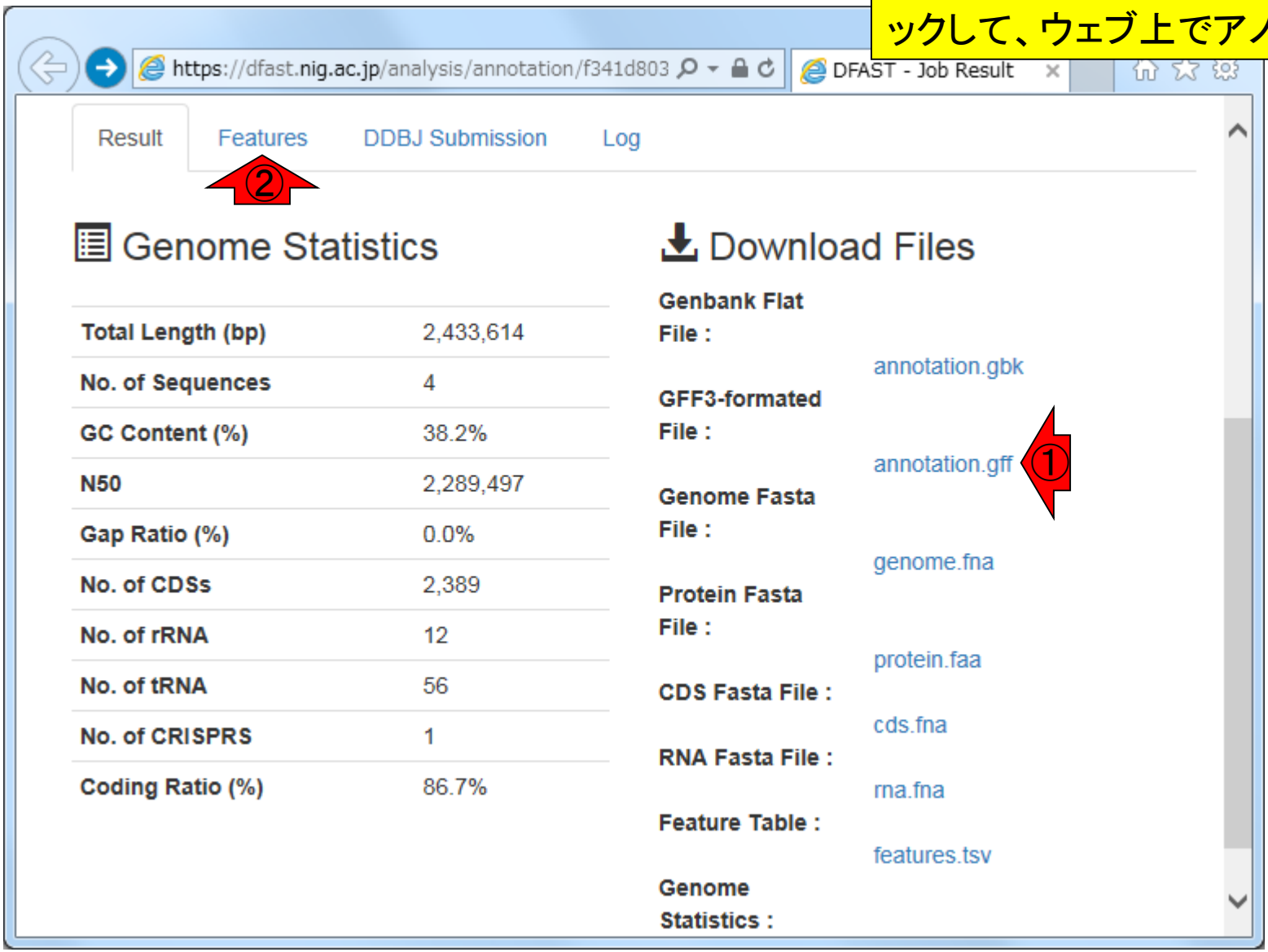
W4-5：結果を眺める

これがDFASTによるアノテーション結果の全体像。①このあたりの数値は、DDBJ Pipeline上でHGAPを実行したときの結果（第7回W9−2）と全く同じで妥当



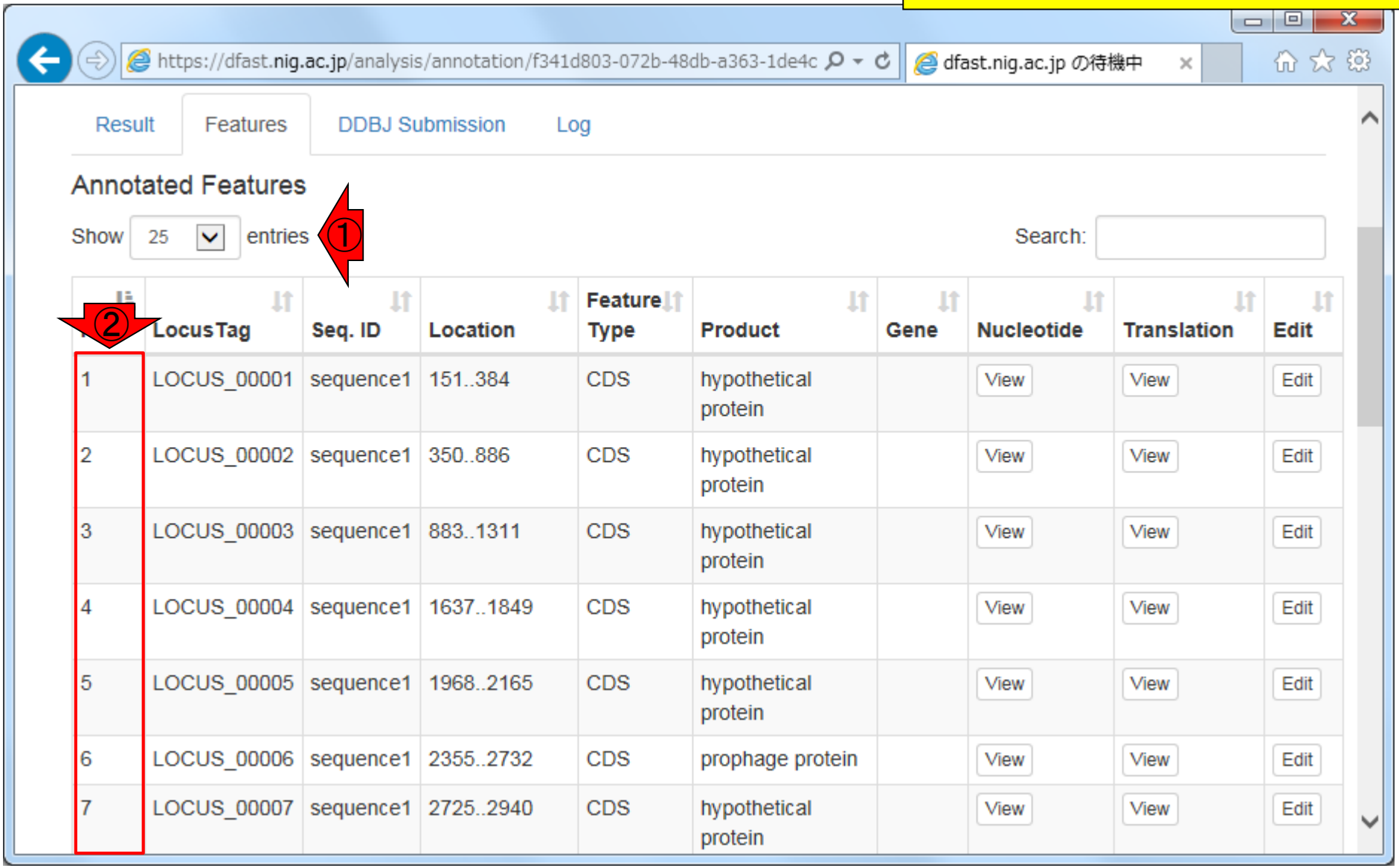
W4-5：結果を眺める

①アノテーションファイルの一般的な形式であるGFFファイルをダウンロードしてエクセルなどで眺めてもよいが、ここでは②Featuresタブをクリックして、ウェブ上でアノテーション結果を眺める



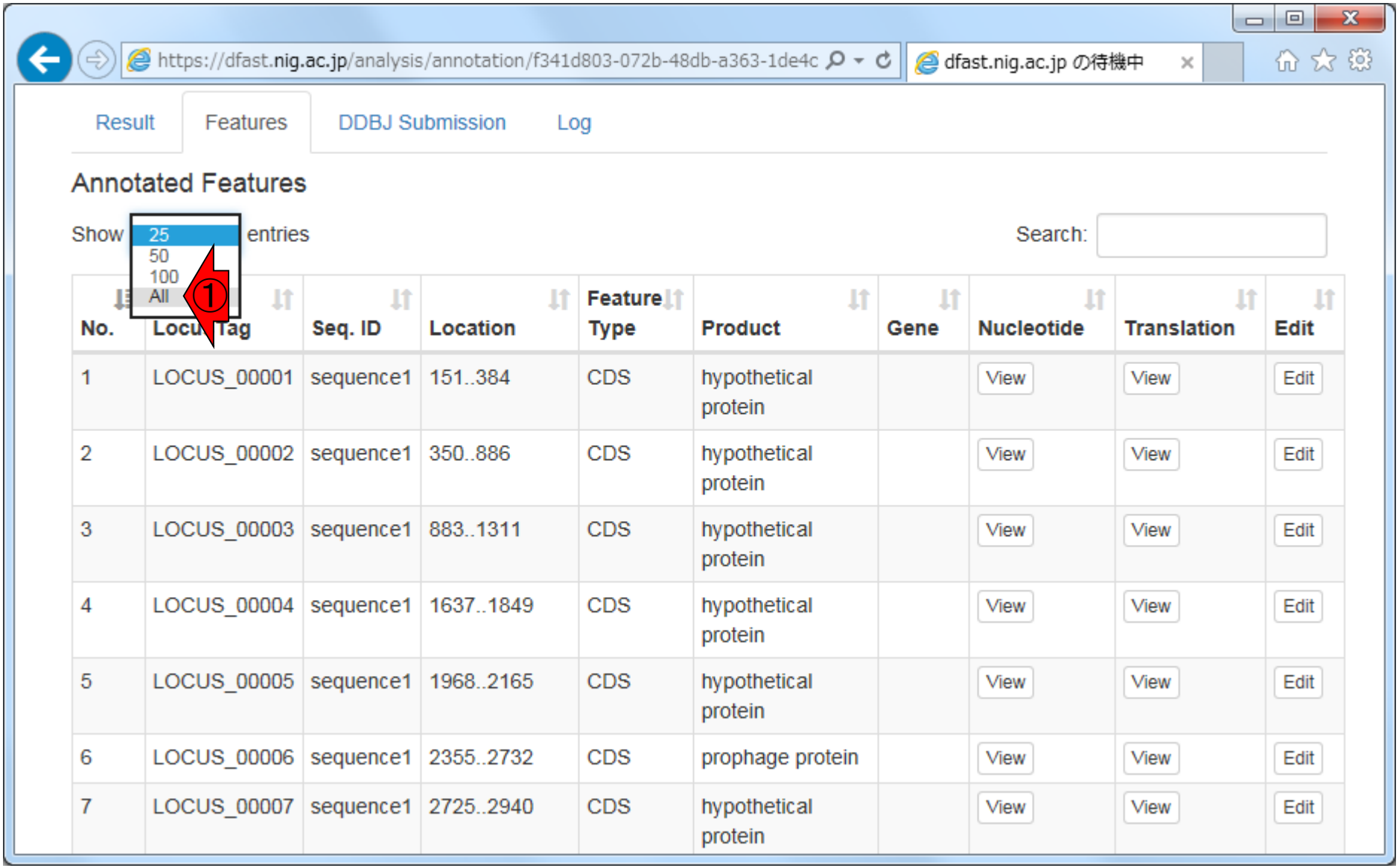
W4-5：結果を眺める

こんな感じになります。①デフォルトはページあたり25エントリーになっているので、②の部分が25行分づつ表示されます



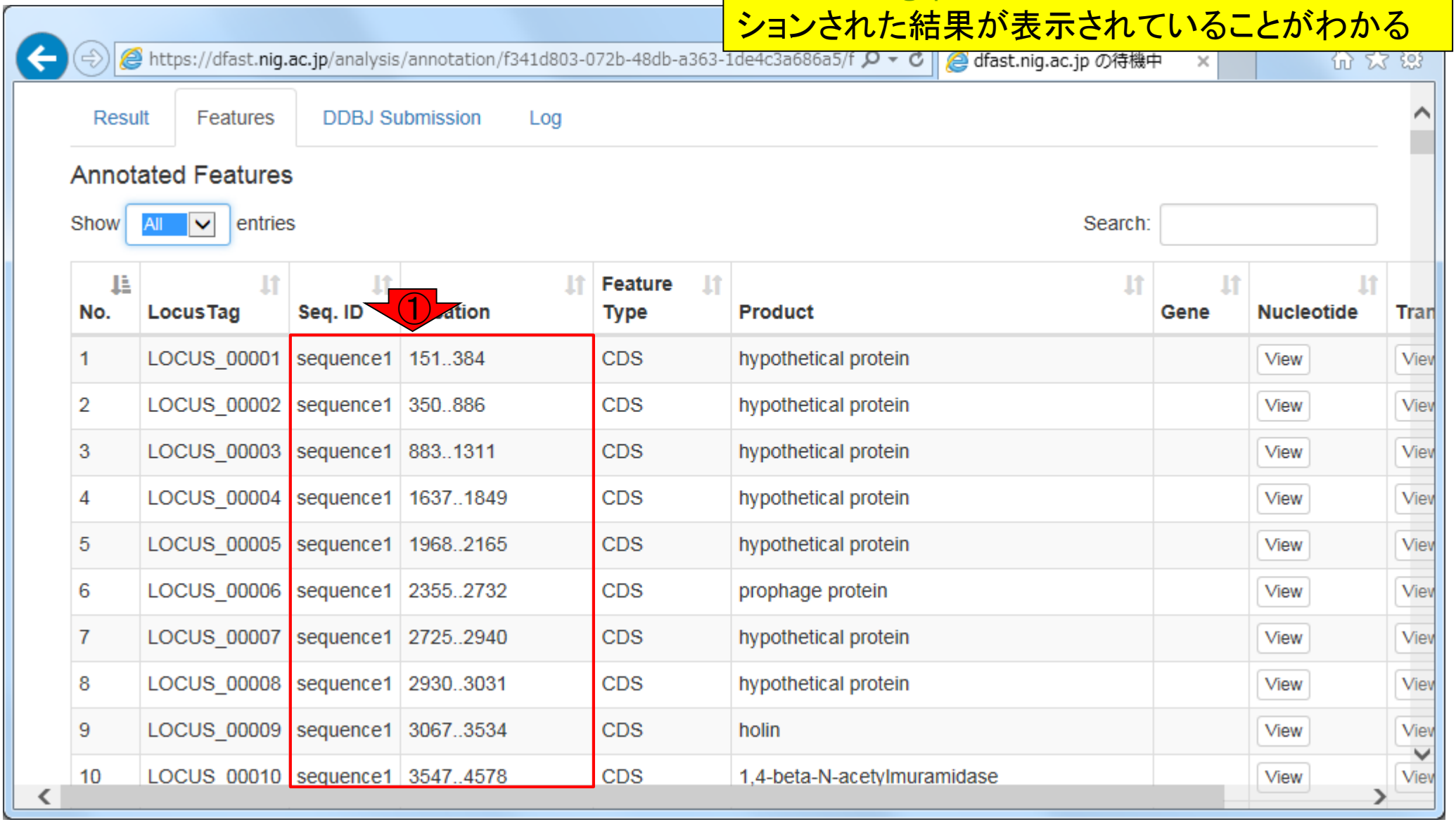
W4-5：結果を眺める

全エントリーを一気に表示させて説明したいので、①Allにする



W4-5：結果を眺める

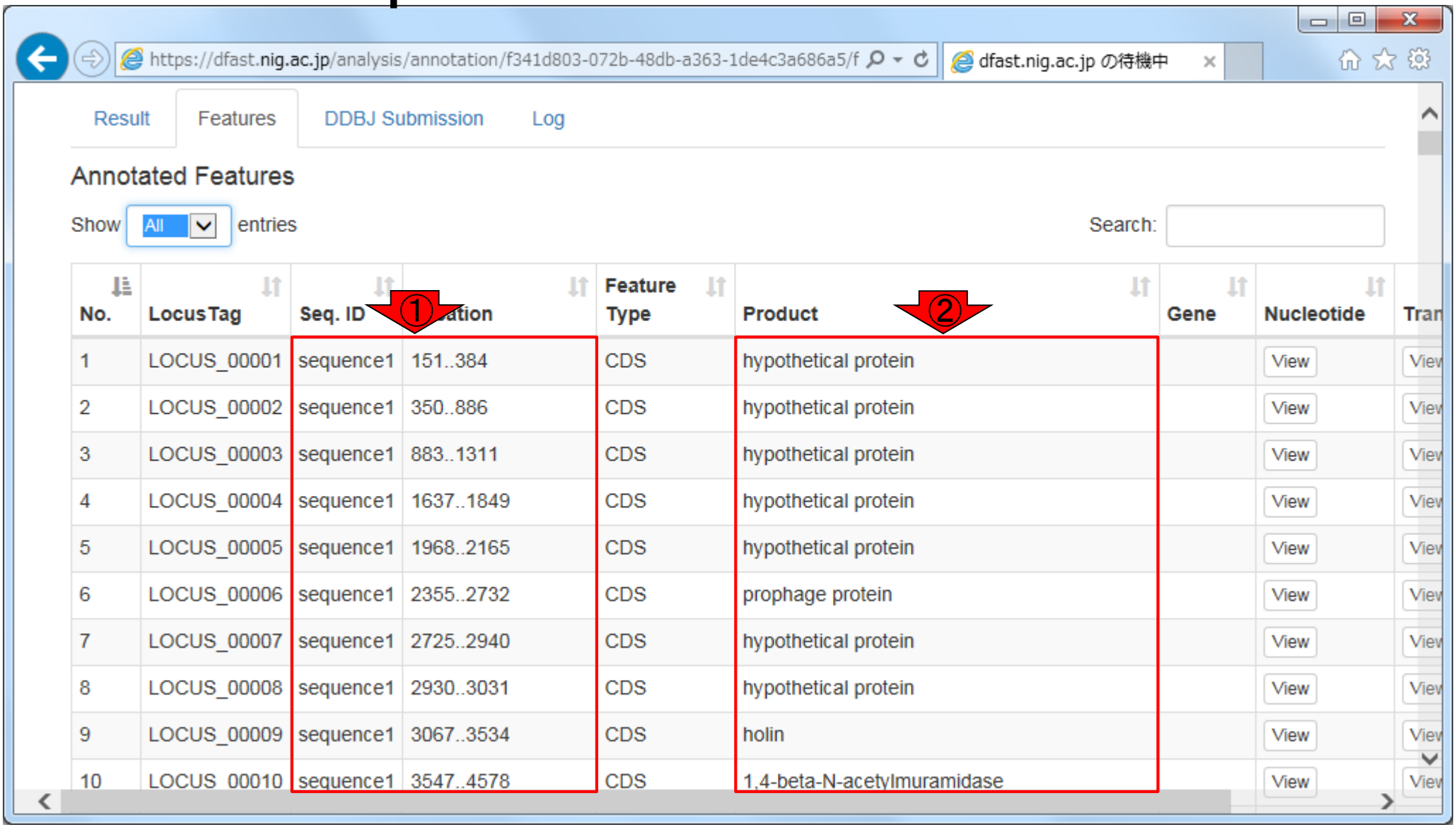
こんな感じになります。横幅が広くなるが細かいことは気にしない。①のあたりを見ることで、入力ファイル(LH_hgap.fa)の配列および座標順にアノテーションされた結果が表示されていることがわかる



No.	Locus Tag	Seq. ID	ation	Feature Type	Product	Gene	Nucleotide	Tran
1	LOCUS_00001	sequence1	151..384	CDS	hypothetical protein		View	View
2	LOCUS_00002	sequence1	350..886	CDS	hypothetical protein		View	View
3	LOCUS_00003	sequence1	883..1311	CDS	hypothetical protein		View	View
4	LOCUS_00004	sequence1	1637..1849	CDS	hypothetical protein		View	View
5	LOCUS_00005	sequence1	1968..2165	CDS	hypothetical protein		View	View
6	LOCUS_00006	sequence1	2355..2732	CDS	prophage protein		View	View
7	LOCUS_00007	sequence1	2725..2940	CDS	hypothetical protein		View	View
8	LOCUS_00008	sequence1	2930..3031	CDS	hypothetical protein		View	View
9	LOCUS_00009	sequence1	3067..3534	CDS	holin		View	View
10	LOCUS_00010	sequence1	3547..4578	CDS	1,4-beta-N-acetylmuramidase		View	View

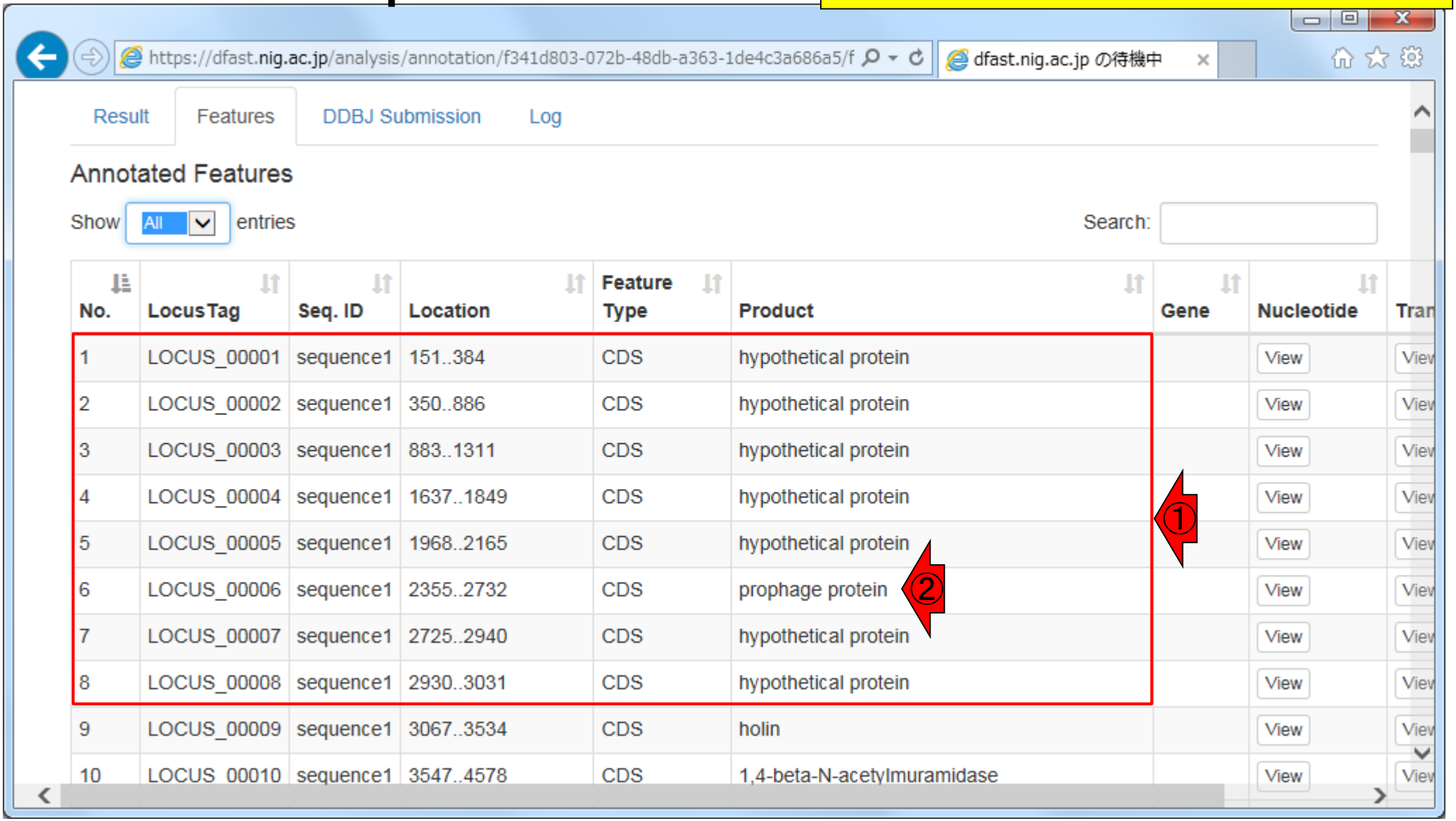
W4-6：sequence1概観

①と②を関連づけながら、アノテーション結果の全体像を把握する



W4-6：sequence1概観

①sequence1の左端（最初の3031 bpまで）は「hypothetical proteinが多いなあ…」とか、②「prophage proteinがある」とか…



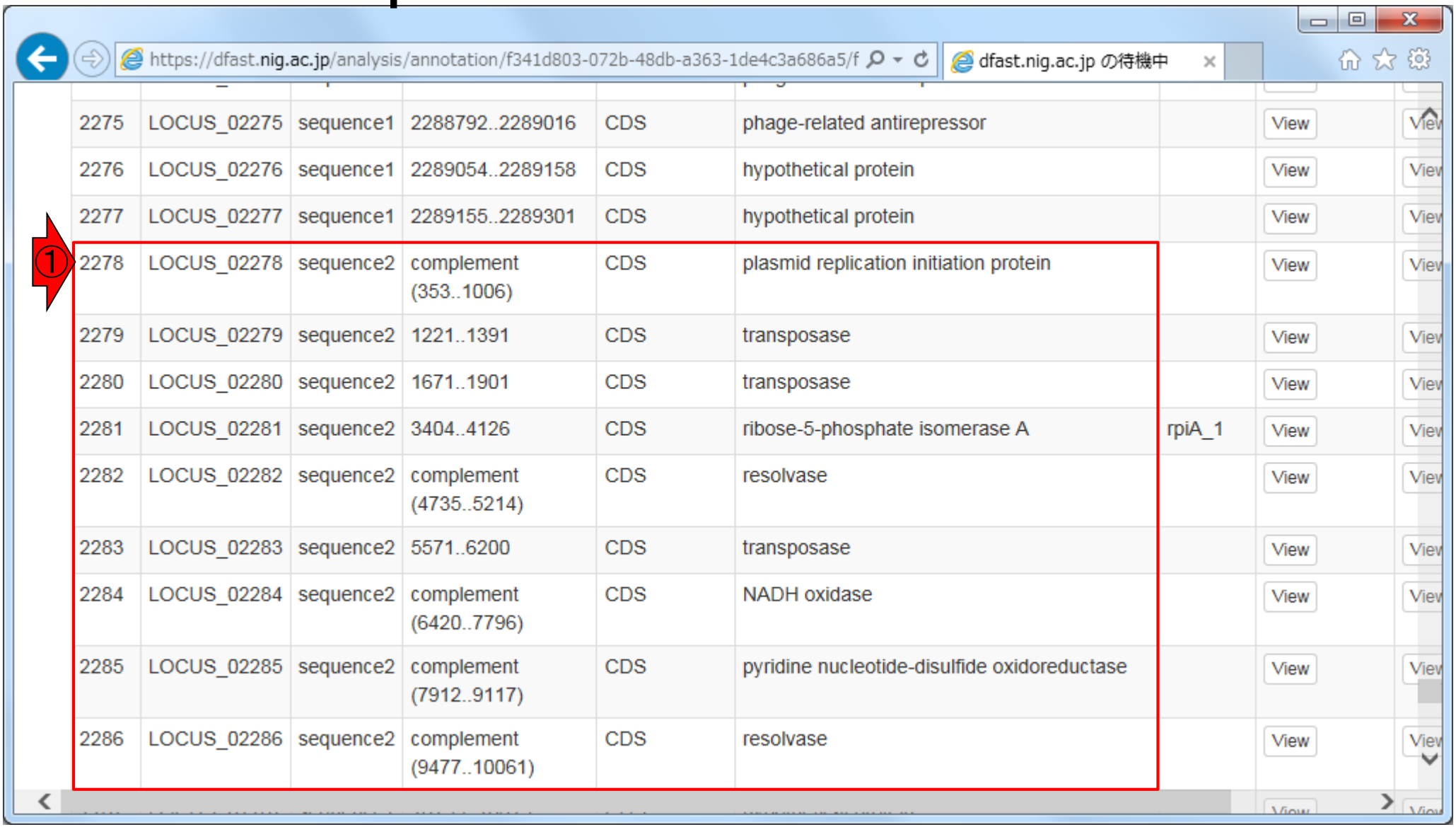
W4-6：sequence1概観

赤枠はsequence1の①19649から②34418 bpの範囲（全部で2,289,497 bpの長さがあるので、このあたりもまだ左端といえる）。赤下線で示すように、このあたりにもファージ(phage)関連のものがちらほら存在

42	LOCUS_00042	sequen	19649..20095	CDS	phage protein		View	View
43	LOCUS_00043	sequence1	20310..20585	CDS	hypothetical protein		View	View
44	LOCUS_00044	sequence1	20578..21057	CDS	hypothetical protein		View	View
45	LOCUS_00045	sequence1	21180..21698	CDS	phage terminase small subunit	xtmA	View	View
46	LOCUS_00046	sequence1	21698..23599	CDS	phage terminase large subunit	xtmB	View	View
47	LOCUS_00047	sequence1	23609..23785	CDS	hypothetical protein		View	View
48	LOCUS_00048	sequence1	23786..24961	CDS	phage portal protein		View	View
49	LOCUS_00049	sequence1	24942..26834	CDS	phage capsid protein		View	View
50	LOCUS_00050	sequence1	27026..27313	CDS	hypothetical protein		View	View
51	LOCUS_00051	sequence1	27294..27671	CDS	hypothetical protein		View	View
52	LOCUS_00052	sequence1	27674..28096	CDS	hypothetical protein		View	View
53	LOCUS_00053	sequence1	28096..28476	CDS	hypothetical protein		View	View
54	LOCUS_00054	sequence1	28480..29088	CDS	phage tail protein		View	View
55	LOCUS_00055	sequence1	29239..29589	CDS	tail protein		View	View
56	LOCUS_00056	sequence1	29793..34418	CDS	phage tail tape measure protein		View	View

W4-7：sequence2概観

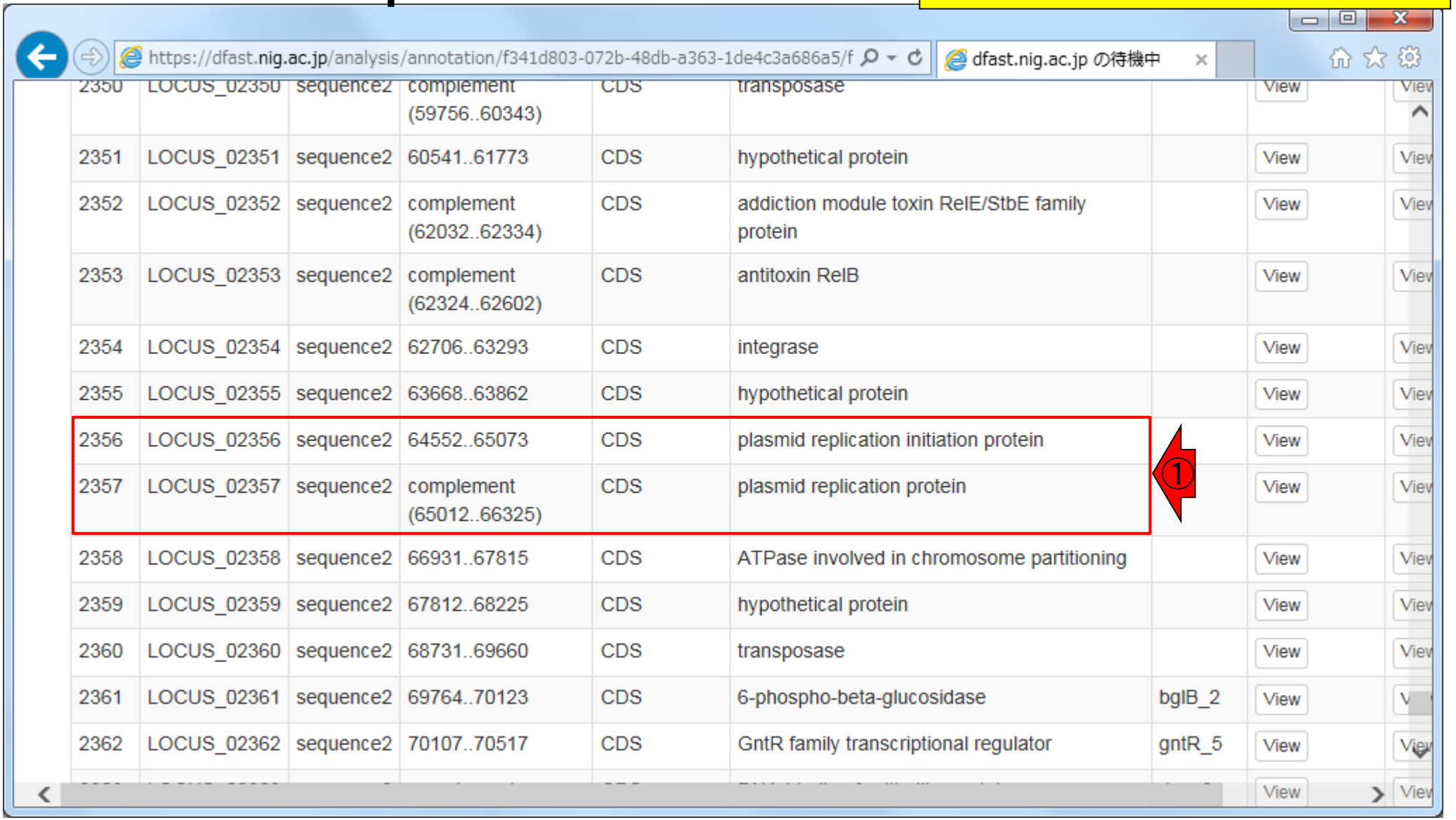
①ここからが、sequence2のアノテーション結果



2275	LOCUS_02275	sequence1	2288792..2289016	CDS	phage-related antirepressor		View	View
2276	LOCUS_02276	sequence1	2289054..2289158	CDS	hypothetical protein		View	View
2277	LOCUS_02277	sequence1	2289155..2289301	CDS	hypothetical protein		View	View
2278	LOCUS_02278	sequence2	complement (353..1006)	CDS	plasmid replication initiation protein		View	View
2279	LOCUS_02279	sequence2	1221..1391	CDS	transposase		View	View
2280	LOCUS_02280	sequence2	1671..1901	CDS	transposase		View	View
2281	LOCUS_02281	sequence2	3404..4126	CDS	ribose-5-phosphate isomerase A	rpiA_1	View	View
2282	LOCUS_02282	sequence2	complement (4735..5214)	CDS	resolvase		View	View
2283	LOCUS_02283	sequence2	5571..6200	CDS	transposase		View	View
2284	LOCUS_02284	sequence2	complement (6420..7796)	CDS	NADH oxidase		View	View
2285	LOCUS_02285	sequence2	complement (7912..9117)	CDS	pyridine nucleotide-disulfide oxidoreductase		View	View
2286	LOCUS_02286	sequence2	complement (9477..10061)	CDS	resolvase		View	View

W4-7：sequence2概観

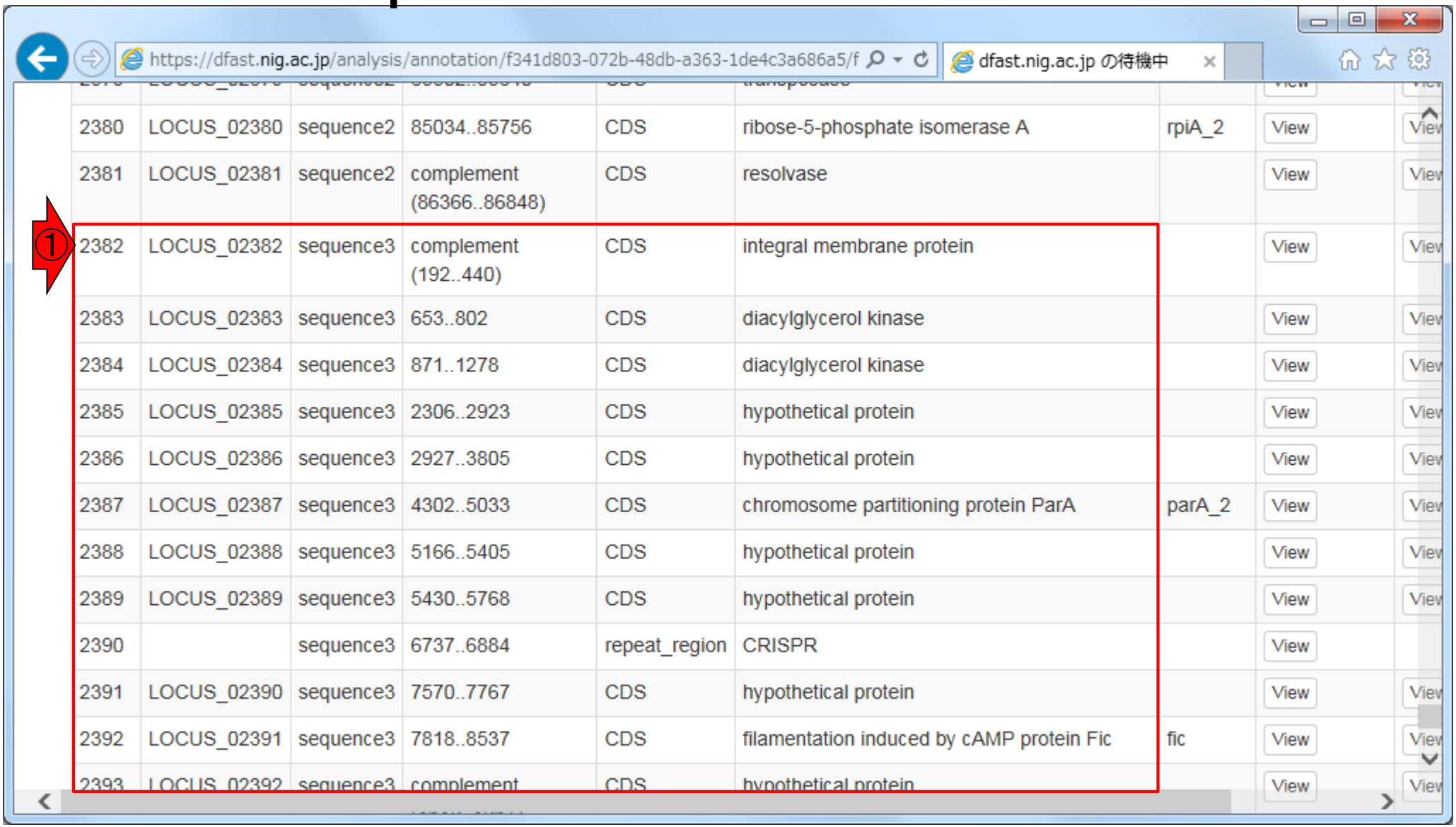
第7回でsequence2は環状のプラスミド配列だろうと予想していたが、①それを補強するアノテーション結果



2350	LOCUS_02350	sequence2	complement (59756..60343)	CDS	transposase		View	View
2351	LOCUS_02351	sequence2	60541..61773	CDS	hypothetical protein		View	View
2352	LOCUS_02352	sequence2	complement (62032..62334)	CDS	addiction module toxin RelE/StbE family protein		View	View
2353	LOCUS_02353	sequence2	complement (62324..62602)	CDS	antitoxin RelB		View	View
2354	LOCUS_02354	sequence2	62706..63293	CDS	integrase		View	View
2355	LOCUS_02355	sequence2	63668..63862	CDS	hypothetical protein		View	View
2356	LOCUS_02356	sequence2	64552..65073	CDS	plasmid replication initiation protein		View	View
2357	LOCUS_02357	sequence2	complement (65012..66325)	CDS	plasmid replication protein		View	View
2358	LOCUS_02358	sequence2	66931..67815	CDS	ATPase involved in chromosome partitioning		View	View
2359	LOCUS_02359	sequence2	67812..68225	CDS	hypothetical protein		View	View
2360	LOCUS_02360	sequence2	68731..69660	CDS	transposase		View	View
2361	LOCUS_02361	sequence2	69764..70123	CDS	6-phospho-beta-glucosidase	bglB_2	View	View
2362	LOCUS_02362	sequence2	70107..70517	CDS	GntR family transcriptional regulator	gntR_5	View	View

W4-8：sequence3概観

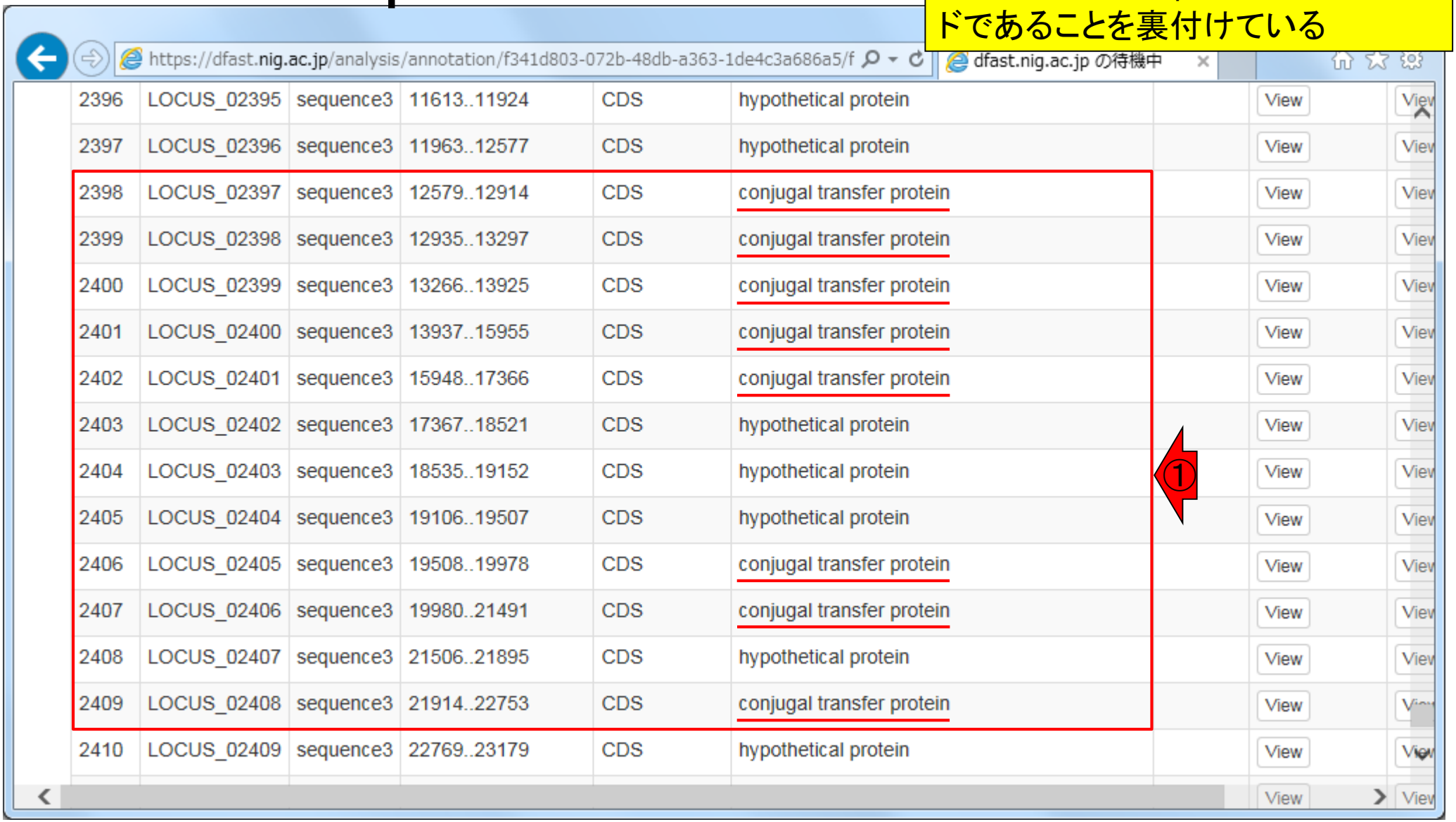
①ここからが、sequence3のアノテーション結果



2380	LOCUS_02380	sequence2	85034..85756	CDS	ribose-5-phosphate isomerase A	rpiA_2	View	View
2381	LOCUS_02381	sequence2	complement (86366..86848)	CDS	resolvase		View	View
① 2382	LOCUS_02382	sequence3	complement (192..440)	CDS	integral membrane protein		View	View
2383	LOCUS_02383	sequence3	653..802	CDS	diacylglycerol kinase		View	View
2384	LOCUS_02384	sequence3	871..1278	CDS	diacylglycerol kinase		View	View
2385	LOCUS_02385	sequence3	2306..2923	CDS	hypothetical protein		View	View
2386	LOCUS_02386	sequence3	2927..3805	CDS	hypothetical protein		View	View
2387	LOCUS_02387	sequence3	4302..5033	CDS	chromosome partitioning protein ParA	parA_2	View	View
2388	LOCUS_02388	sequence3	5166..5405	CDS	hypothetical protein		View	View
2389	LOCUS_02389	sequence3	5430..5768	CDS	hypothetical protein		View	View
2390		sequence3	6737..6884	repeat_region	CRISPR		View	
2391	LOCUS_02390	sequence3	7570..7767	CDS	hypothetical protein		View	View
2392	LOCUS_02391	sequence3	7818..8537	CDS	filamentation induced by cAMP protein Fic	fic	View	View
2393	LOCUS_02392	sequence3	complement	CDS	hypothetical protein		View	View

W4-8：sequence3概観

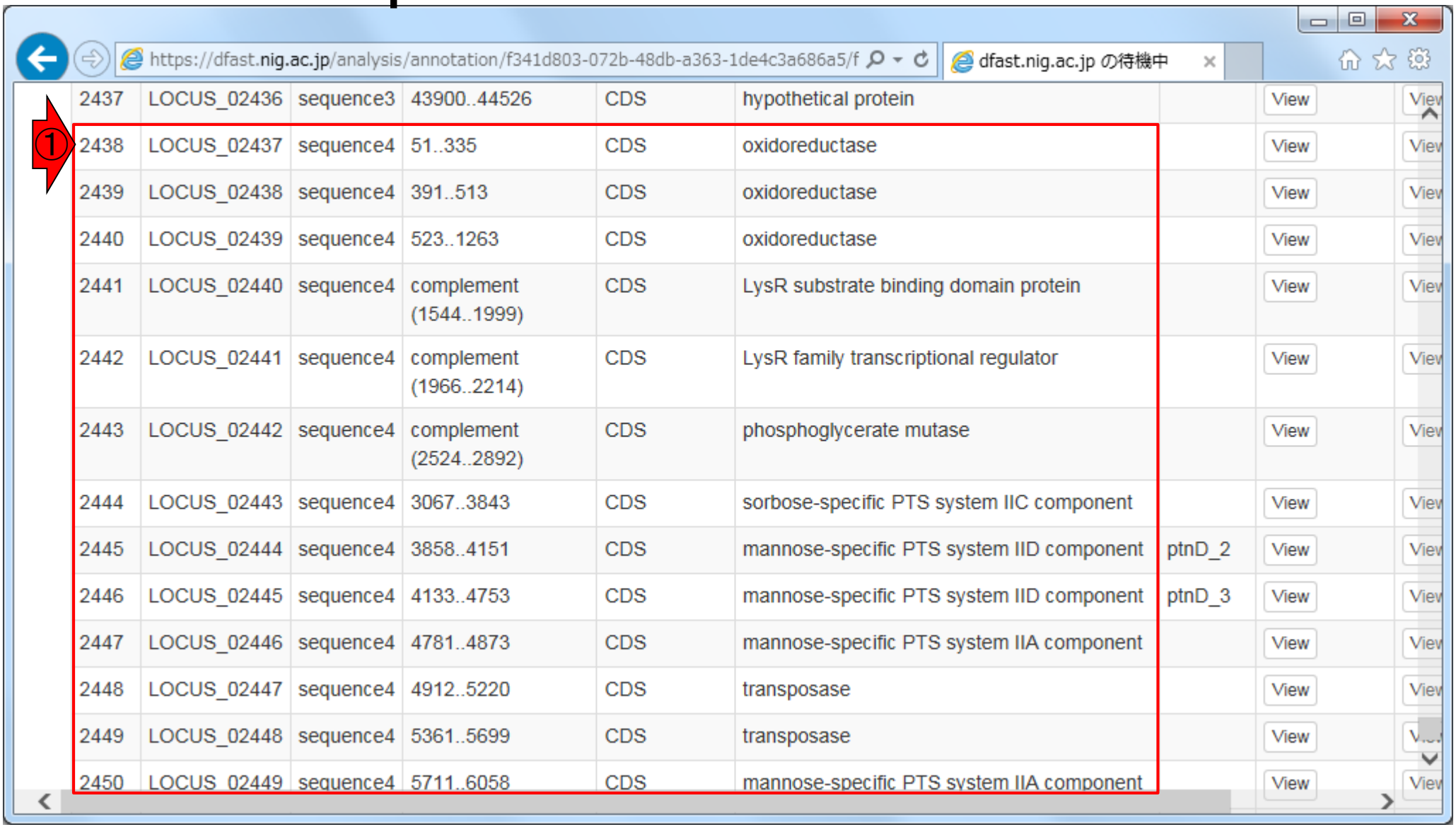
①赤下線で示す接合伝達(conjugal transfer)関連遺伝子が多く見られることからも、sequence3がプラスミドであることを裏付けている



2396	LOCUS_02395	sequence3	11613..11924	CDS	hypothetical protein	View	View
2397	LOCUS_02396	sequence3	11963..12577	CDS	hypothetical protein	View	View
2398	LOCUS_02397	sequence3	12579..12914	CDS	conjugal transfer protein	View	View
2399	LOCUS_02398	sequence3	12935..13297	CDS	conjugal transfer protein	View	View
2400	LOCUS_02399	sequence3	13266..13925	CDS	conjugal transfer protein	View	View
2401	LOCUS_02400	sequence3	13937..15955	CDS	conjugal transfer protein	View	View
2402	LOCUS_02401	sequence3	15948..17366	CDS	conjugal transfer protein	View	View
2403	LOCUS_02402	sequence3	17367..18521	CDS	hypothetical protein	View	View
2404	LOCUS_02403	sequence3	18535..19152	CDS	hypothetical protein	View	View
2405	LOCUS_02404	sequence3	19106..19507	CDS	hypothetical protein	View	View
2406	LOCUS_02405	sequence3	19508..19978	CDS	conjugal transfer protein	View	View
2407	LOCUS_02406	sequence3	19980..21491	CDS	conjugal transfer protein	View	View
2408	LOCUS_02407	sequence3	21506..21895	CDS	hypothetical protein	View	View
2409	LOCUS_02408	sequence3	21914..22753	CDS	conjugal transfer protein	View	View
2410	LOCUS_02409	sequence3	22769..23179	CDS	hypothetical protein	View	View

W4-9：sequence4概観

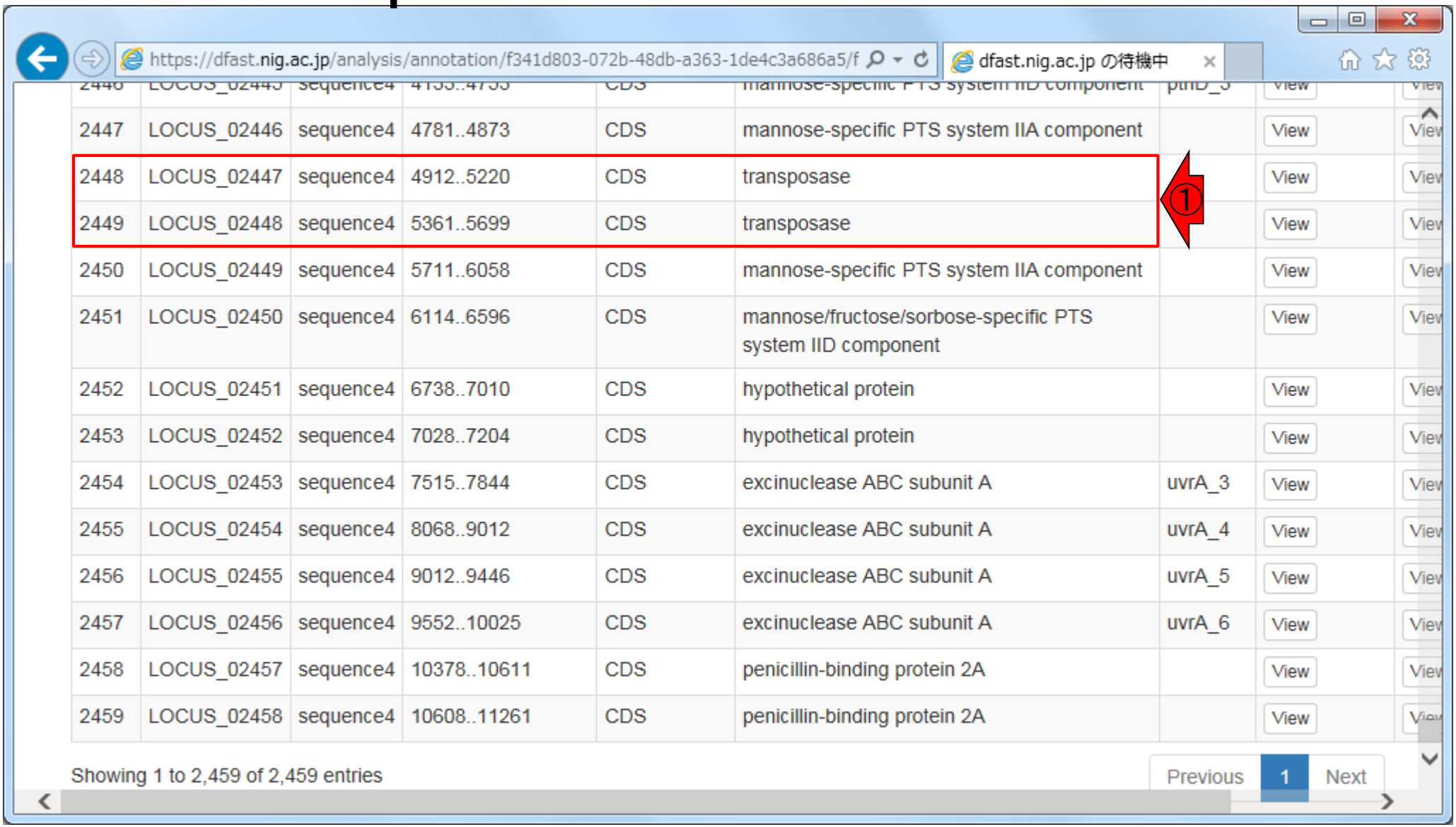
①ここからが、sequence4のアノテーション結果



2437	LOCUS_02436	sequence3	43900..44526	CDS	hypothetical protein		View	View
2438	LOCUS_02437	sequence4	51..335	CDS	oxidoreductase		View	View
2439	LOCUS_02438	sequence4	391..513	CDS	oxidoreductase		View	View
2440	LOCUS_02439	sequence4	523..1263	CDS	oxidoreductase		View	View
2441	LOCUS_02440	sequence4	complement (1544..1999)	CDS	LysR substrate binding domain protein		View	View
2442	LOCUS_02441	sequence4	complement (1966..2214)	CDS	LysR family transcriptional regulator		View	View
2443	LOCUS_02442	sequence4	complement (2524..2892)	CDS	phosphoglycerate mutase		View	View
2444	LOCUS_02443	sequence4	3067..3843	CDS	sorbose-specific PTS system IIC component		View	View
2445	LOCUS_02444	sequence4	3858..4151	CDS	mannose-specific PTS system IID component	ptnD_2	View	View
2446	LOCUS_02445	sequence4	4133..4753	CDS	mannose-specific PTS system IID component	ptnD_3	View	View
2447	LOCUS_02446	sequence4	4781..4873	CDS	mannose-specific PTS system IIA component		View	View
2448	LOCUS_02447	sequence4	4912..5220	CDS	transposase		View	View
2449	LOCUS_02448	sequence4	5361..5699	CDS	transposase		View	View
2450	LOCUS_02449	sequence4	5711..6058	CDS	mannose-specific PTS system IIA component		View	View

①ここにtransposaseがあります

W4-9：sequence4概観



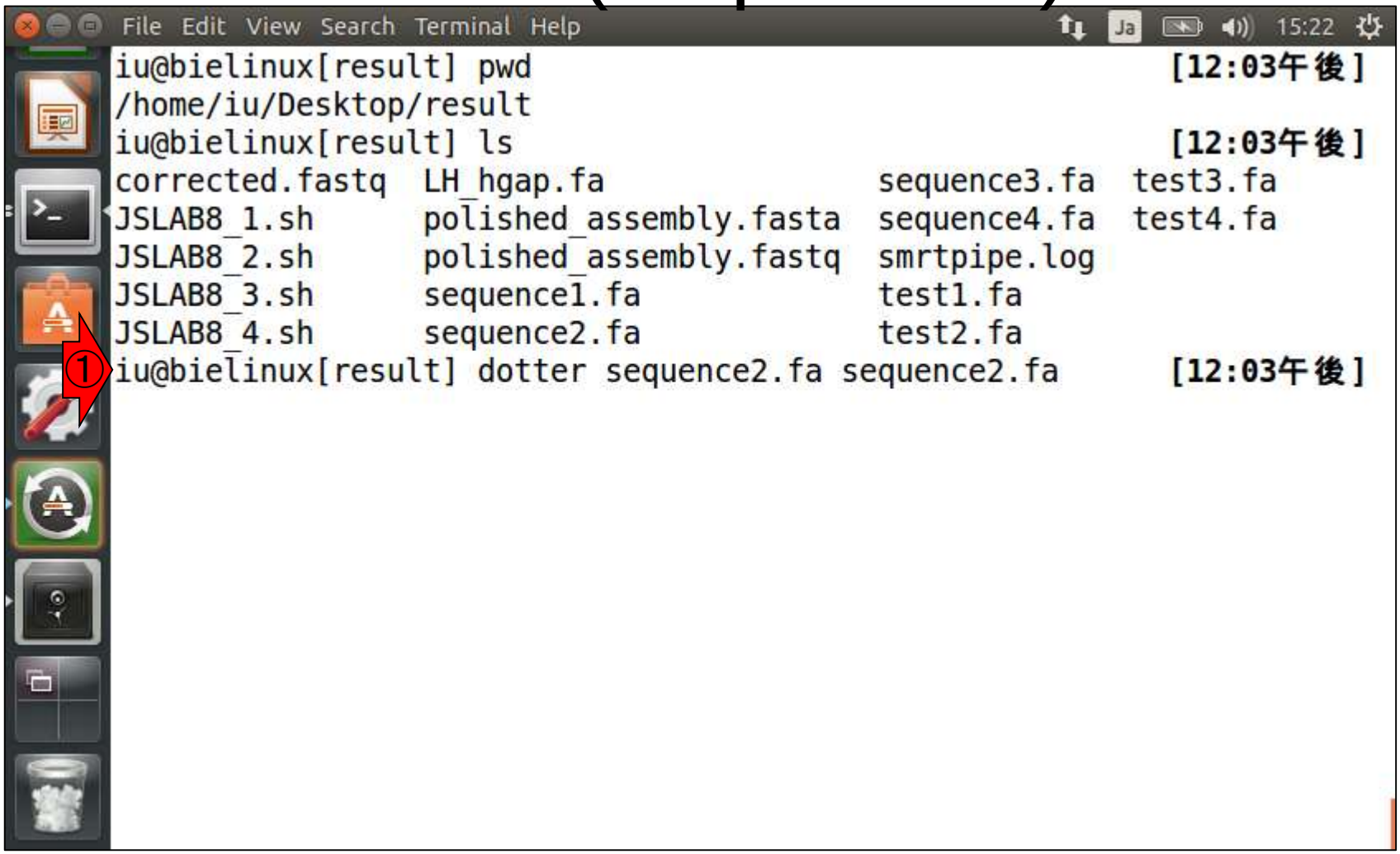
2447	LOCUS_02446	sequence4	4781..4873	CDS	mannose-specific PTS system IIA component		View	View
2448	LOCUS_02447	sequence4	4912..5220	CDS	transposase		View	View
2449	LOCUS_02448	sequence4	5361..5699	CDS	transposase		View	View
2450	LOCUS_02449	sequence4	5711..6058	CDS	mannose-specific PTS system IIA component		View	View
2451	LOCUS_02450	sequence4	6114..6596	CDS	mannose/fructose/sorbose-specific PTS system IID component		View	View
2452	LOCUS_02451	sequence4	6738..7010	CDS	hypothetical protein		View	View
2453	LOCUS_02452	sequence4	7028..7204	CDS	hypothetical protein		View	View
2454	LOCUS_02453	sequence4	7515..7844	CDS	excinuclease ABC subunit A	uvrA_3	View	View
2455	LOCUS_02454	sequence4	8068..9012	CDS	excinuclease ABC subunit A	uvrA_4	View	View
2456	LOCUS_02455	sequence4	9012..9446	CDS	excinuclease ABC subunit A	uvrA_5	View	View
2457	LOCUS_02456	sequence4	9552..10025	CDS	excinuclease ABC subunit A	uvrA_6	View	View
2458	LOCUS_02457	sequence4	10378..10611	CDS	penicillin-binding protein 2A		View	View
2459	LOCUS_02458	sequence4	10608..11261	CDS	penicillin-binding protein 2A		View	View

Showing 1 to 2,459 of 2,459 entries

Previous 1 Next

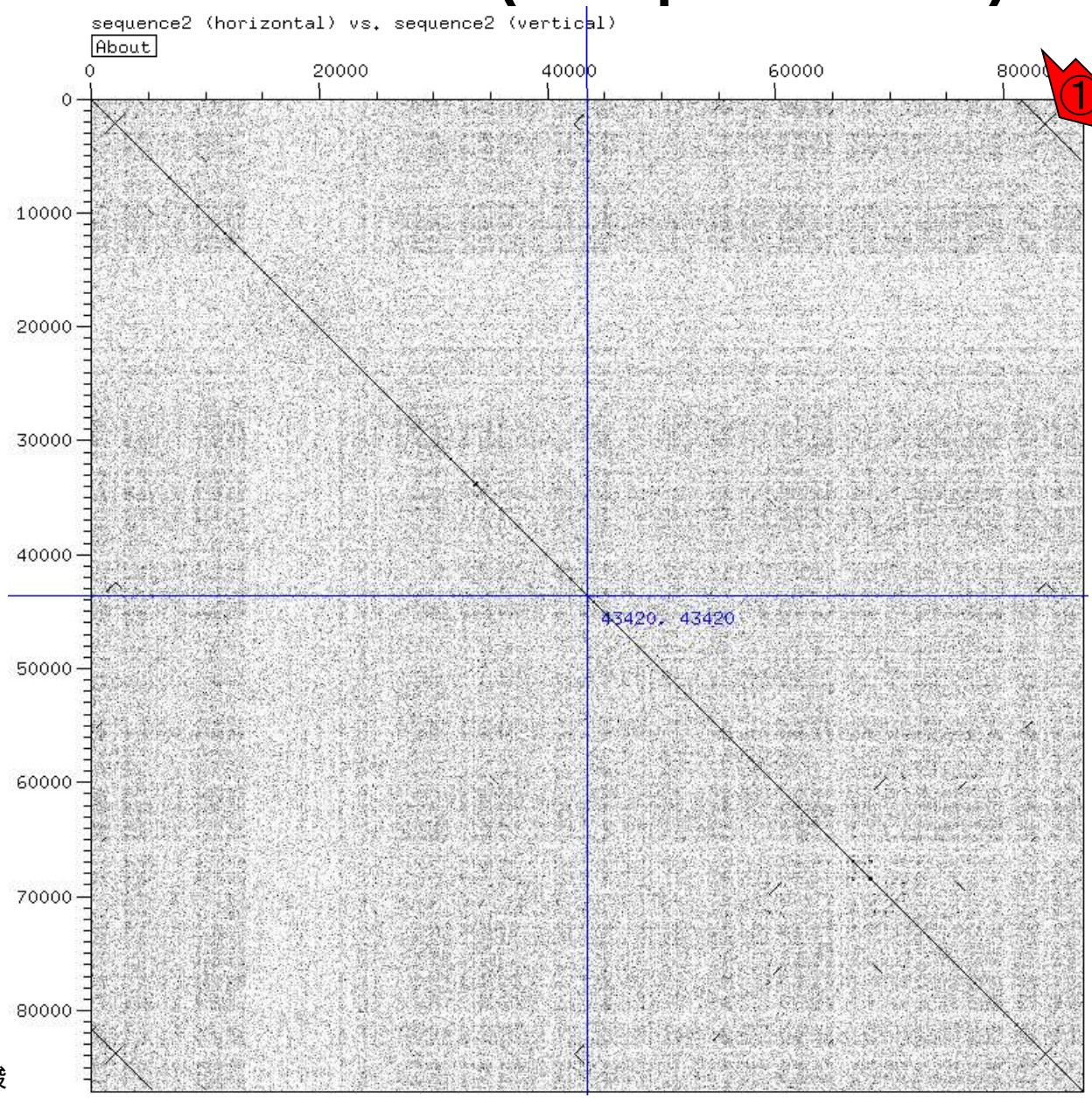
①dotterでsequence2同士を比較

W5-1：dotter (sequence2)



Sonnhammer and Durbin, *Gene*, **167**: GC1-10, 1995

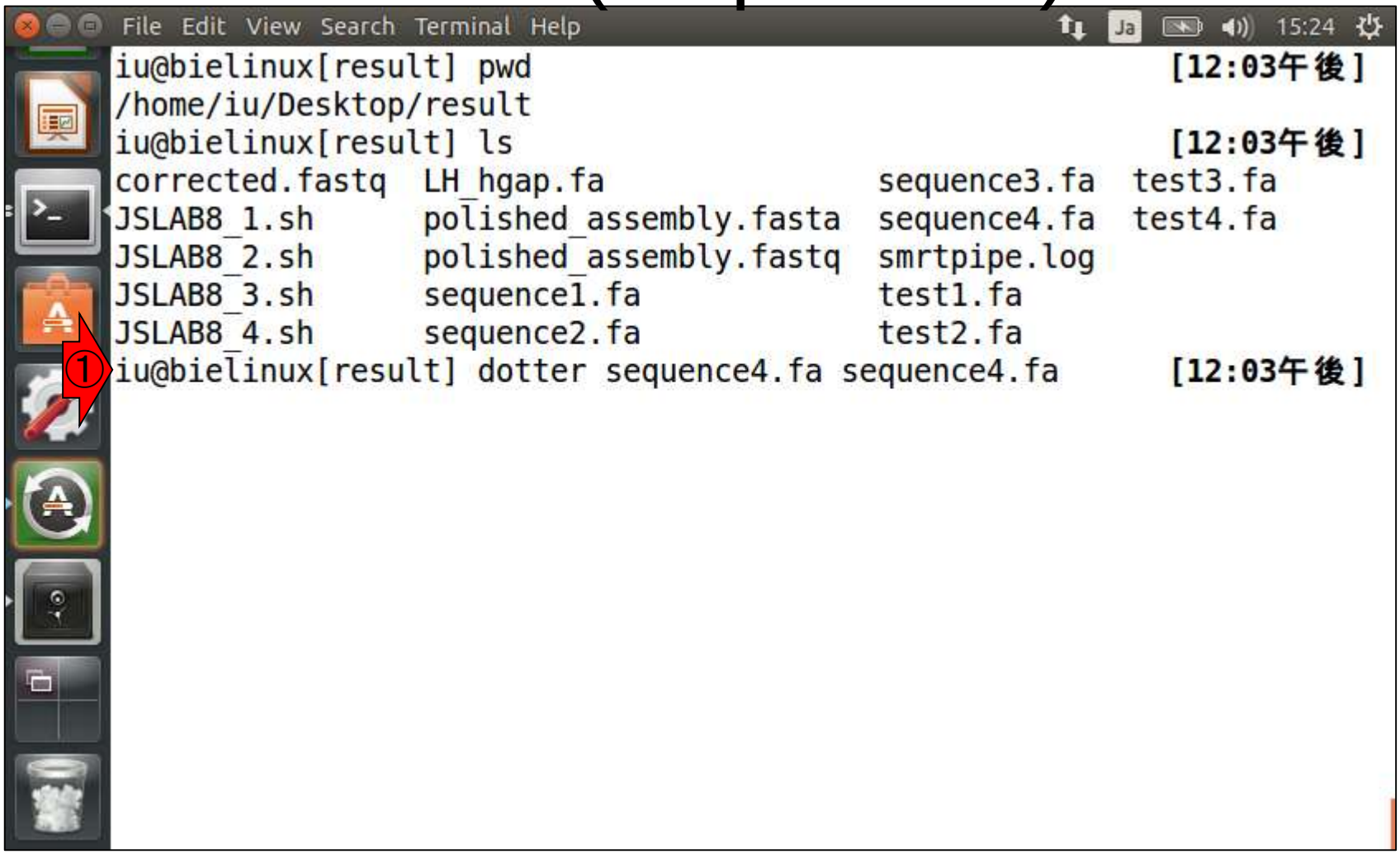
W5-1：dotter (sequence2)



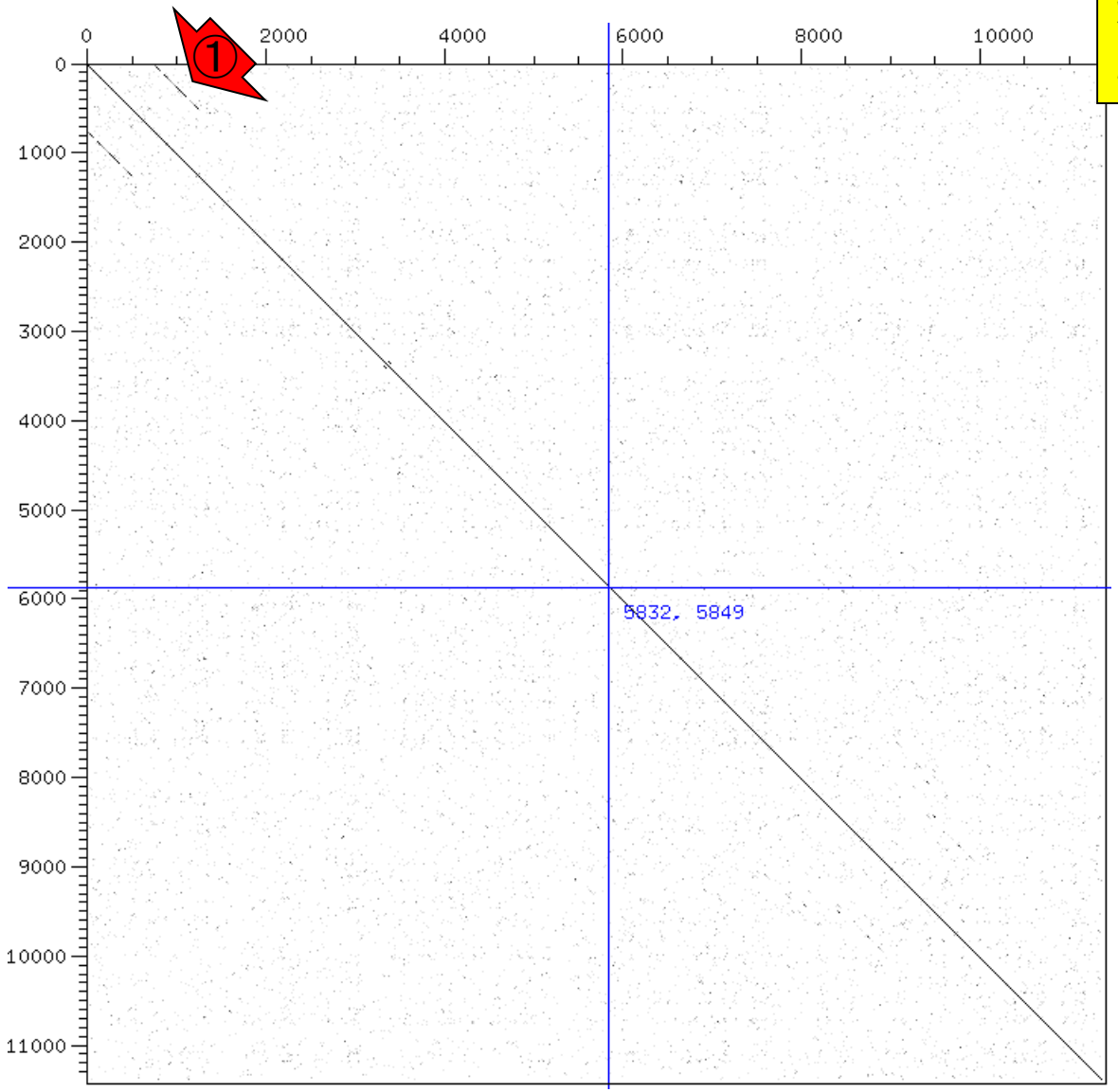
実行結果。①このあたりで5,000 bpほど重複していることがわかる。両末端の5,000 bpあたりはアノテーション結果もほぼ同じなのだろうと妄想する

①dotterでsequence4同士を比較

W5-2：dotter (sequence4)



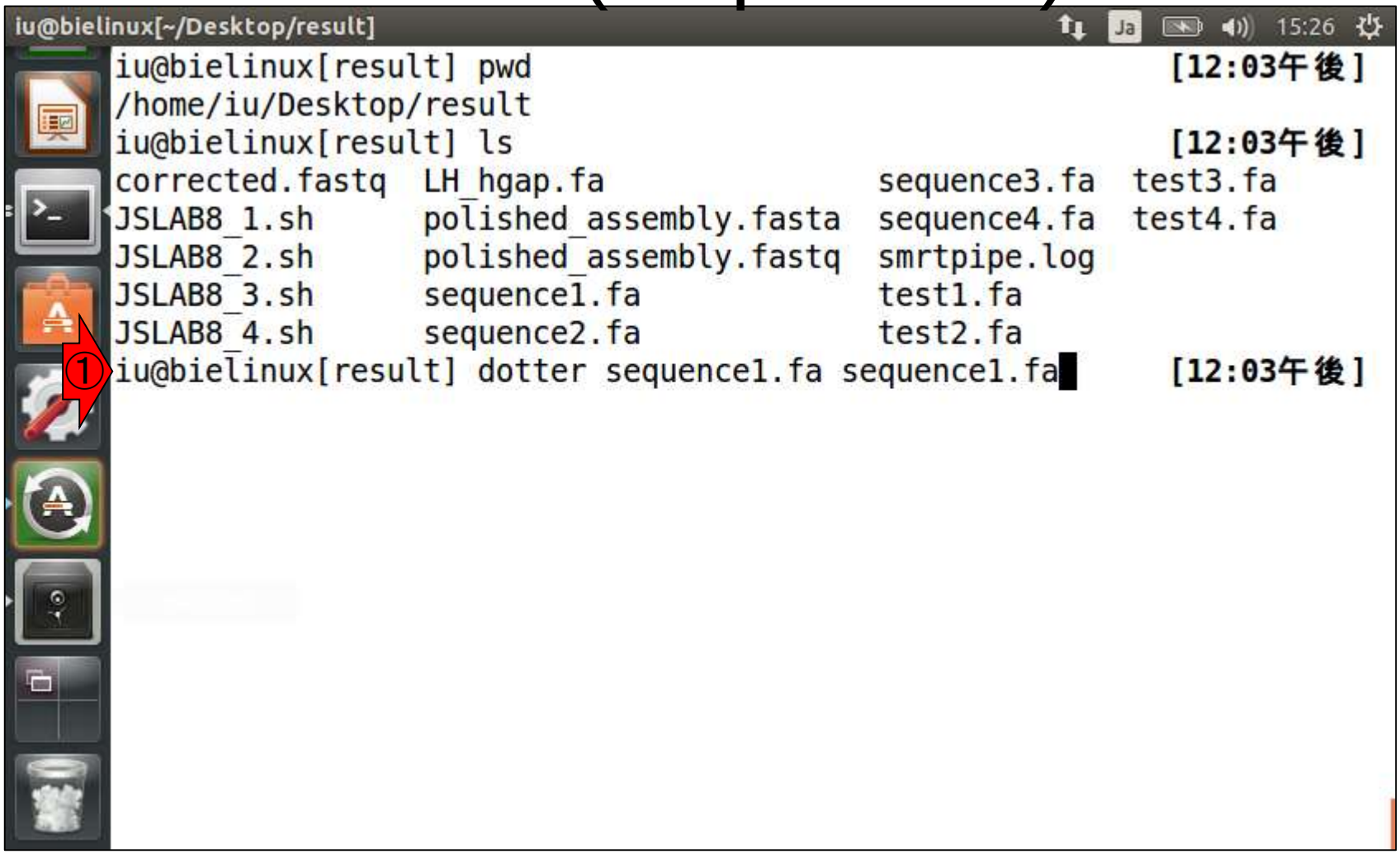
W5-2：dotter (sequence4)



①この結果は、(1 – 500)番目と(750 – 1,350)番目付近の領域が似ていることを意味する。しかし両末端ではないので、sequence4は環状（プラスミド）ではないのだろうと判断

①dotterでsequence1同士を比較

W5-3：dotter (sequence1)



W5-3：dotter (sequence1)

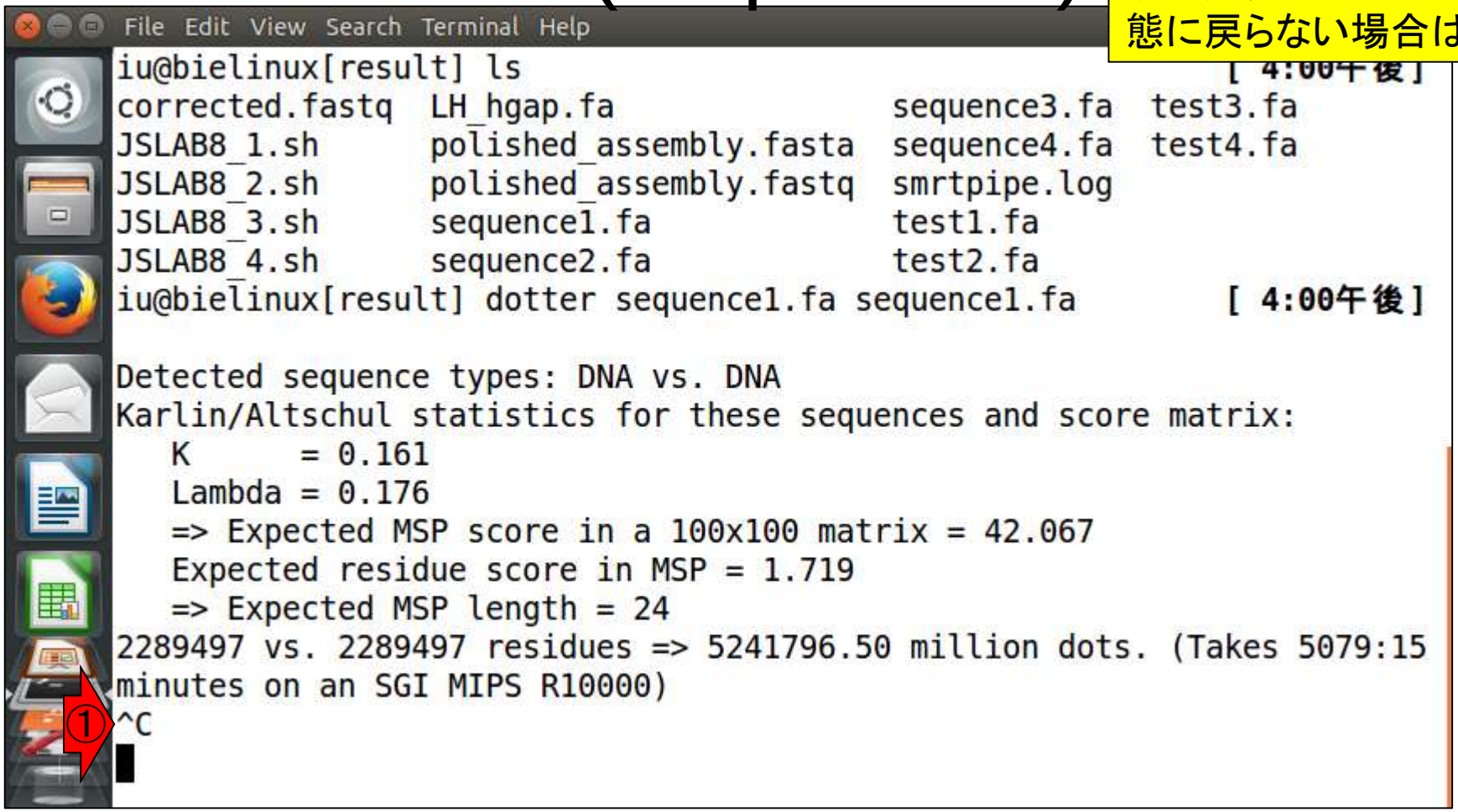
この状態からいつまでも変わらないので、「①CTRL + ②C」で脱出

```
File  Edit  View  Search  Terminal  Help                    Ja  16:01
/home/iu/Desktop/result
iu@bielinux[result] ls                                        [ 4:00午後]
corrected.fastq  LH_hgap.fa                 sequence3.fa  test3.fa
JSLAB8_1.sh      polished_assembly.fasta    sequence4.fa  test4.fa
JSLAB8_2.sh      polished_assembly.fastq    smrtpipe.log
JSLAB8_3.sh      sequence1.fa               test1.fa
JSLAB8_4.sh      sequence2.fa               test2.fa
iu@bielinux[result] dotter sequence1.fa sequence1.fa          [ 4:00午後]

Detected sequence types: DNA vs.
Karlin/Altschul statistics for th
   K      = 0.161
   Lambda = 0.176
   => Expected MSP score in a 100
   Expected residue score in MSP
   => Expected MSP length = 24
2289497 vs. 2289497 residues => 5
minutes on an SGI MIPS R10000)
```

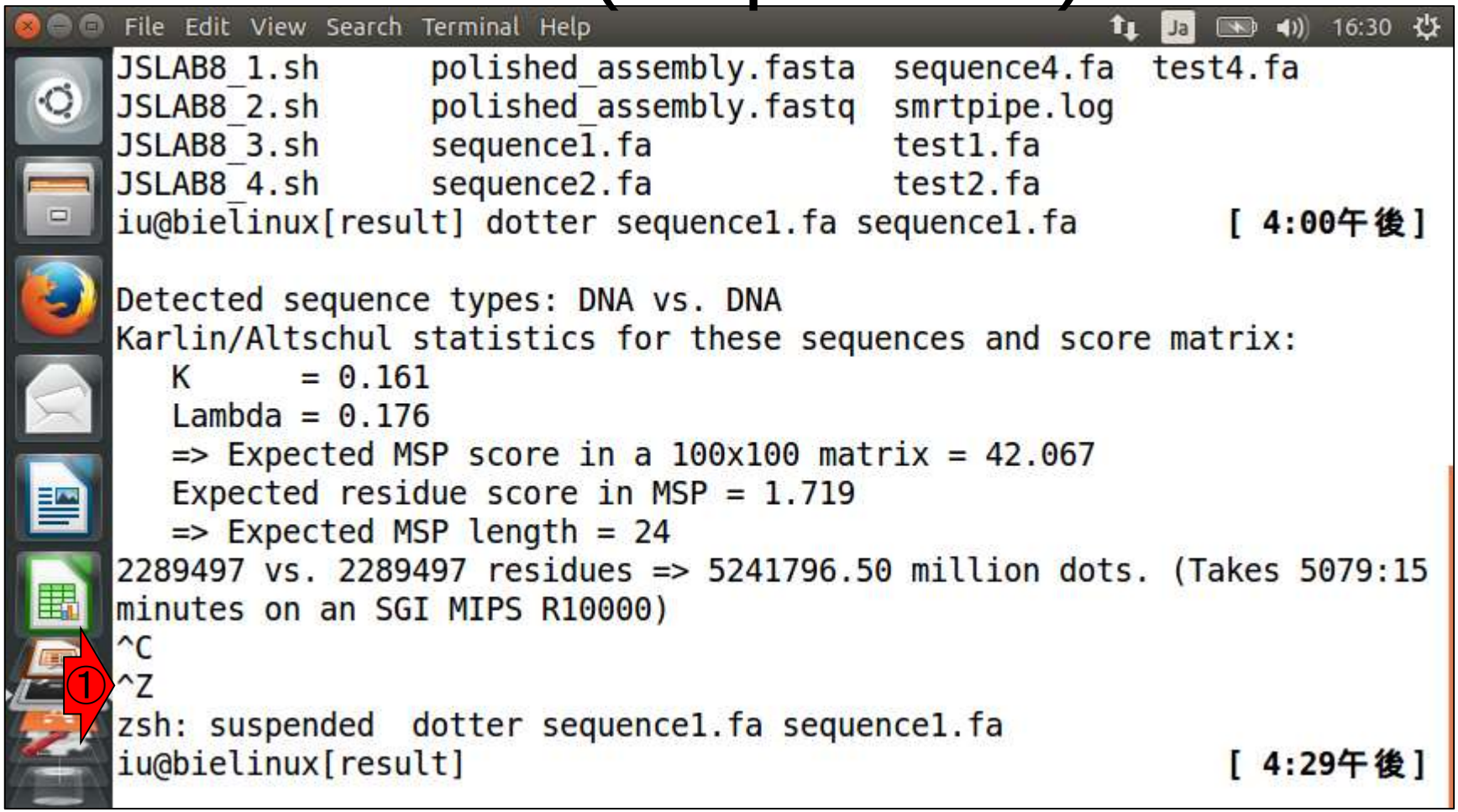
W5-3：dotter (sequence1)

①「CTRL + C」を押しても変化がない場合は、リターンキーを押したりする。それでもコマンド入力待ち状態に戻らない場合は、「CTRL + Z」

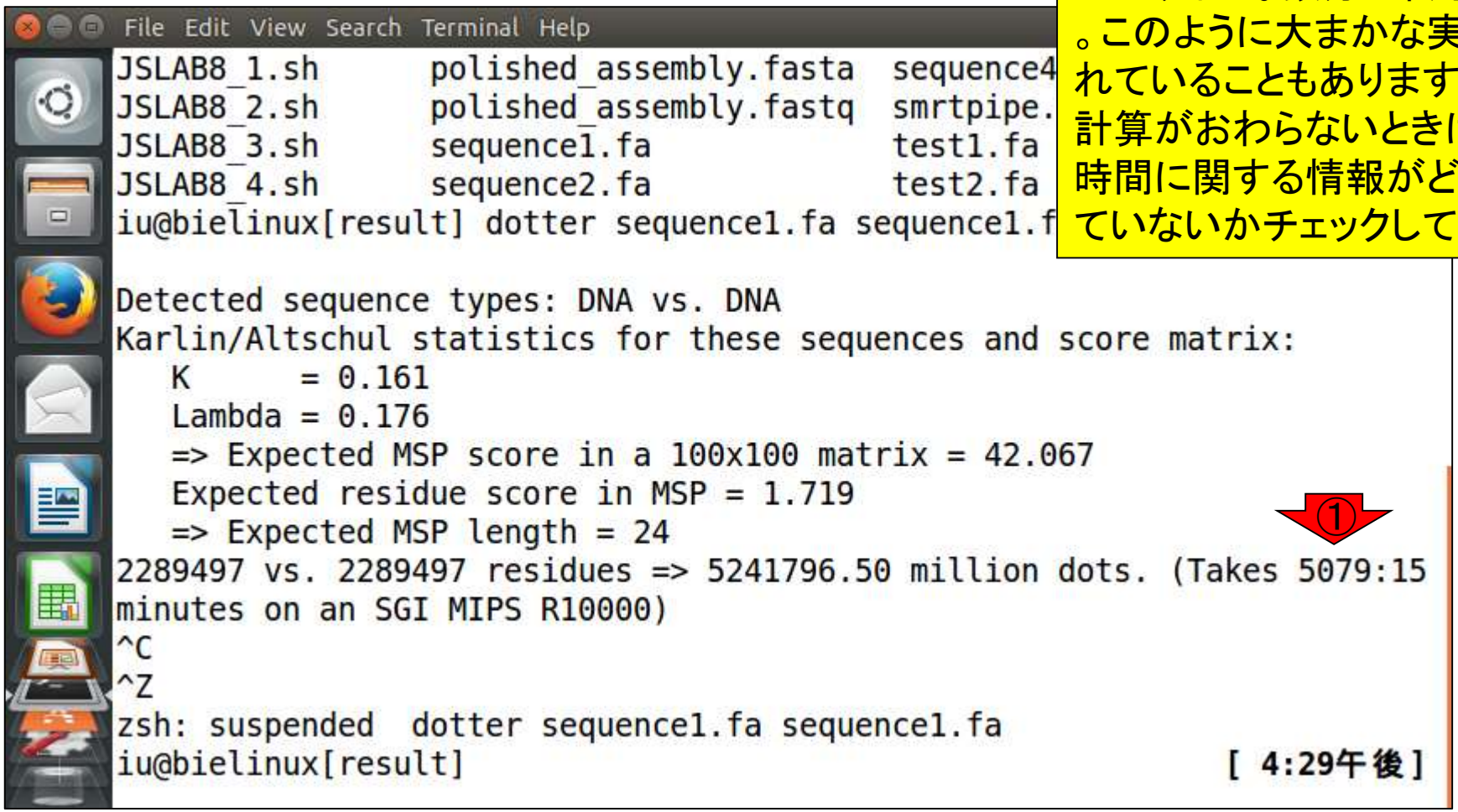


W5-3：dotter (sequence1)

①「CTRL + Z」だと、コマンド入力待ち状態に戻ることができます



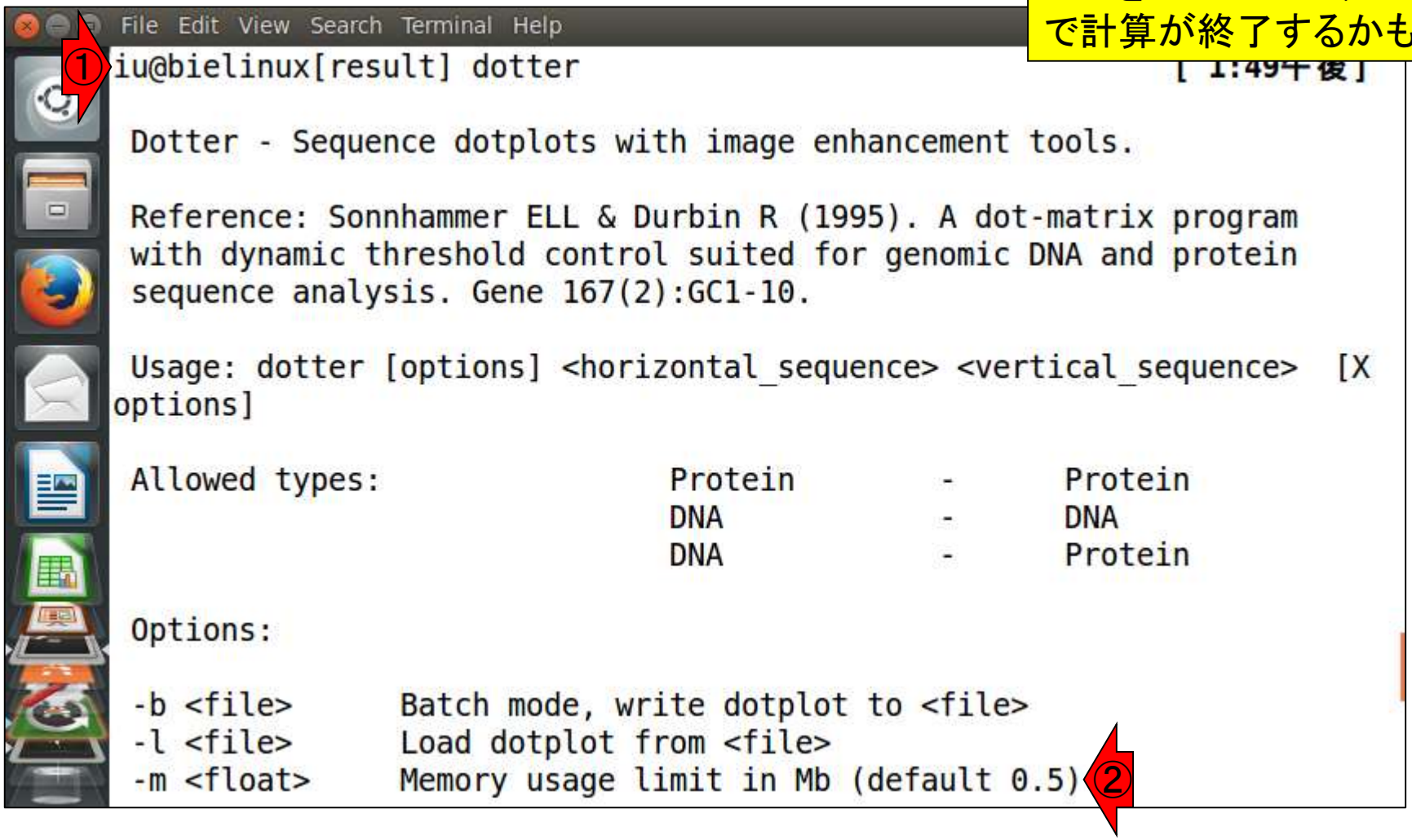
W5-4：dotterのオプション



ちなみに①実行時間は5079分と書かれています。84.65時間なので、約3.5日ですねw。数分で終了しないわけだ。このように大まかな実行時間が書かれていることもありますので、なかなか計算がおわらないときは、①のような時間に関する情報がどこかに書かれていないかチェックしてもいいでしょう

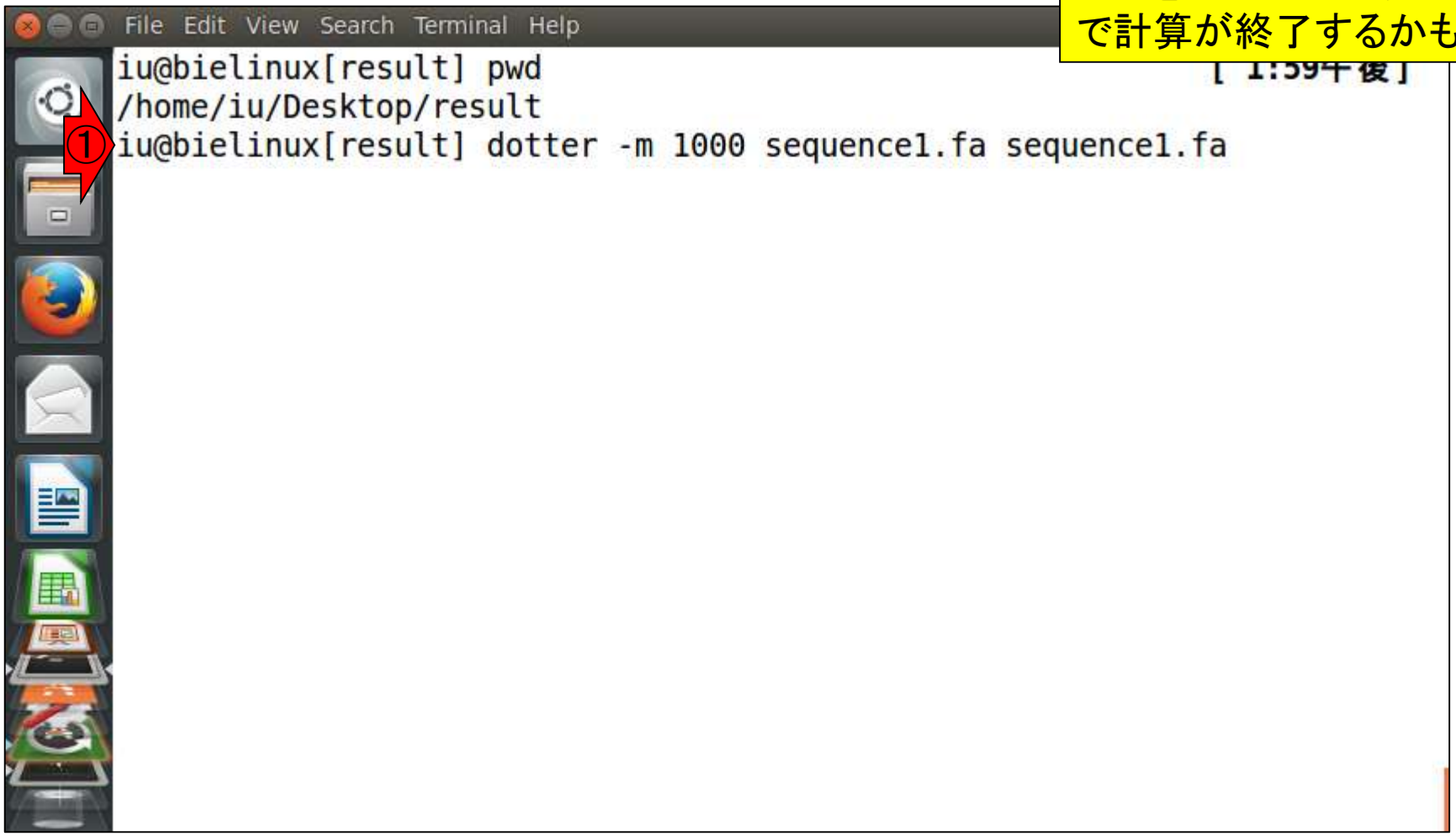
W5-4：dotterのオプション

①でdotterのオプションを眺めると、②デフォルトの使用メモリはたった0.5Mb!これを1000とかにすれば10時間程度で計算が終了するかもしれません。



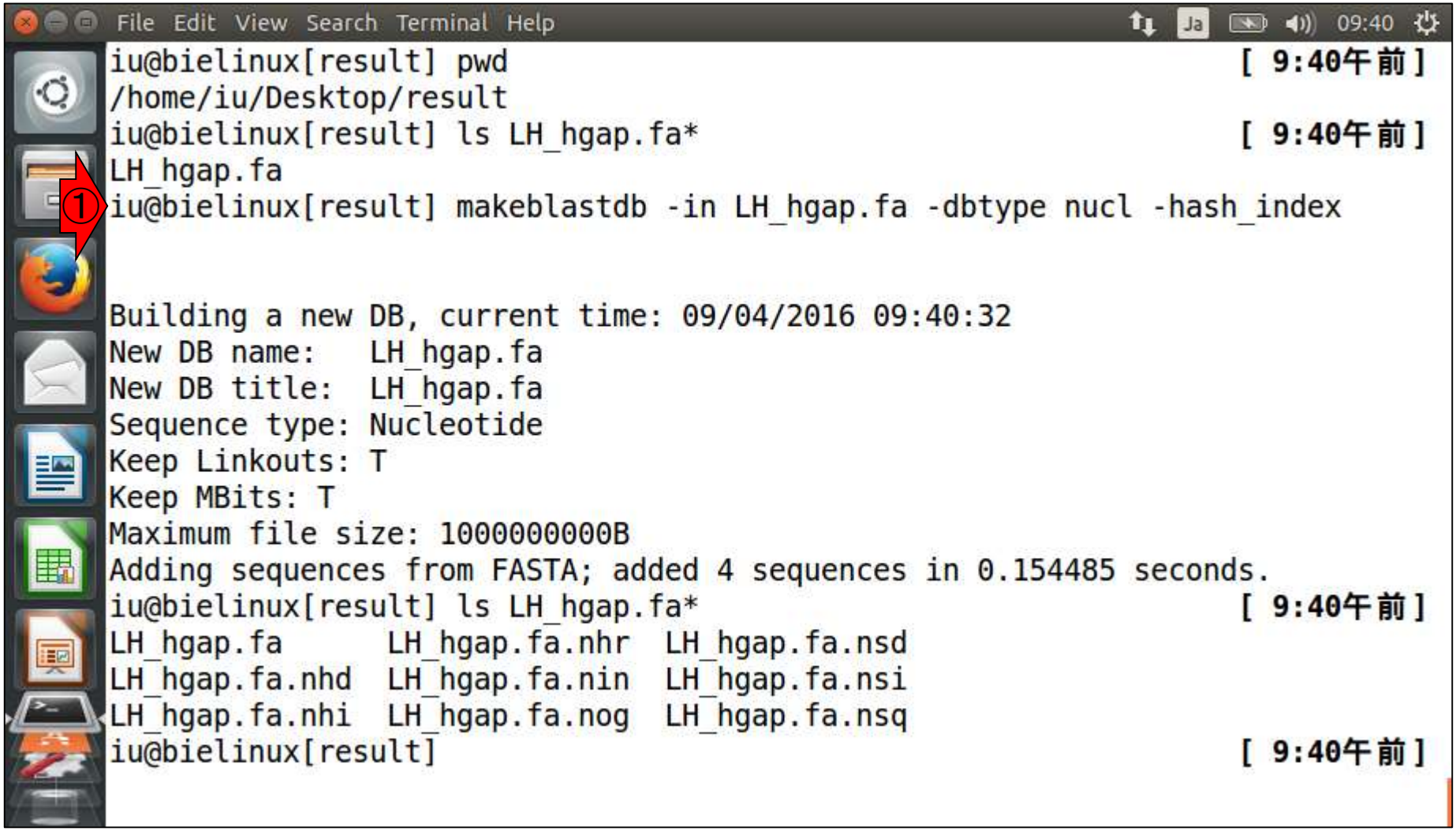
W5-4：dotterのオプション

①でdotterのオプションを眺めると、②デフォルトの使用メモリはたった0.5Mb!これを1000とかにすれば10時間程度で計算が終了するかもしれません。



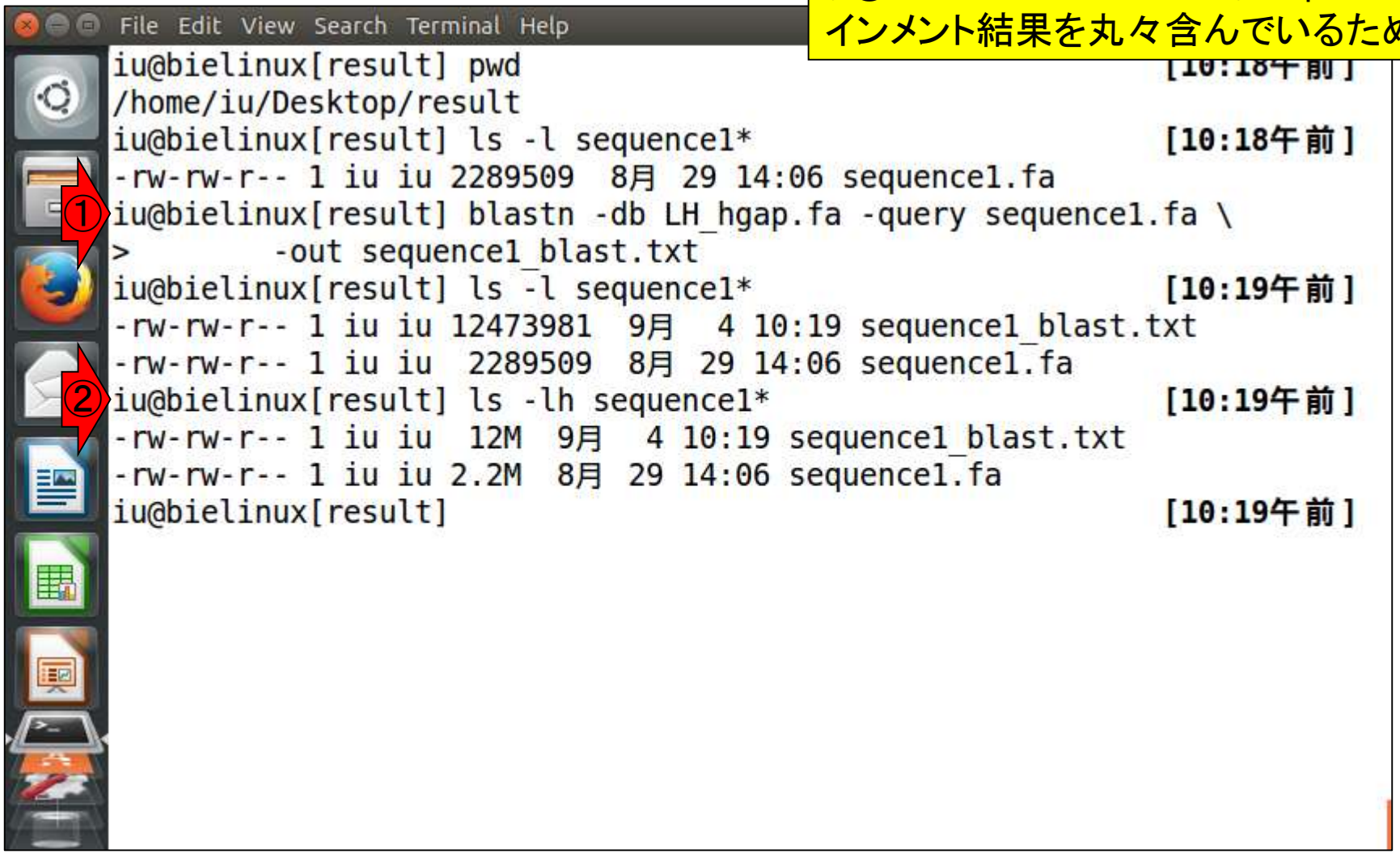
W6-1：makeblastdb

①makeblastdbの実行。第7回W15-1と基本的に同じ



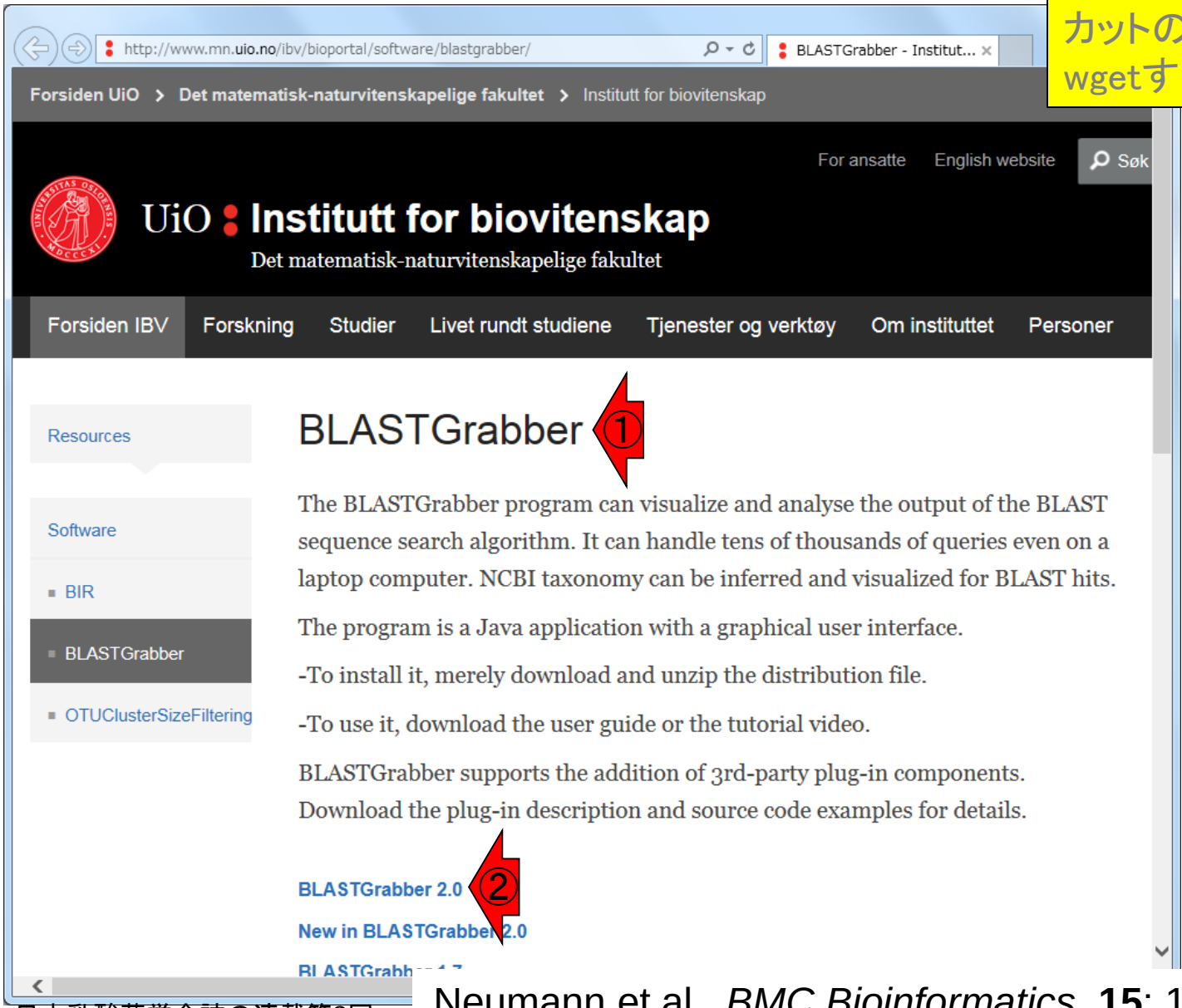
W6-2：blastn

①DB側をLH_hgap.fa、query側をsequence1.fa、出力ファイルをsequence1_blast.txtとしてblastnを実行。②出力ファイルは12MB。sequence1.fa同士のアラインメント結果を丸々含んでいるためであろう



W7-1：BLASTGrabber

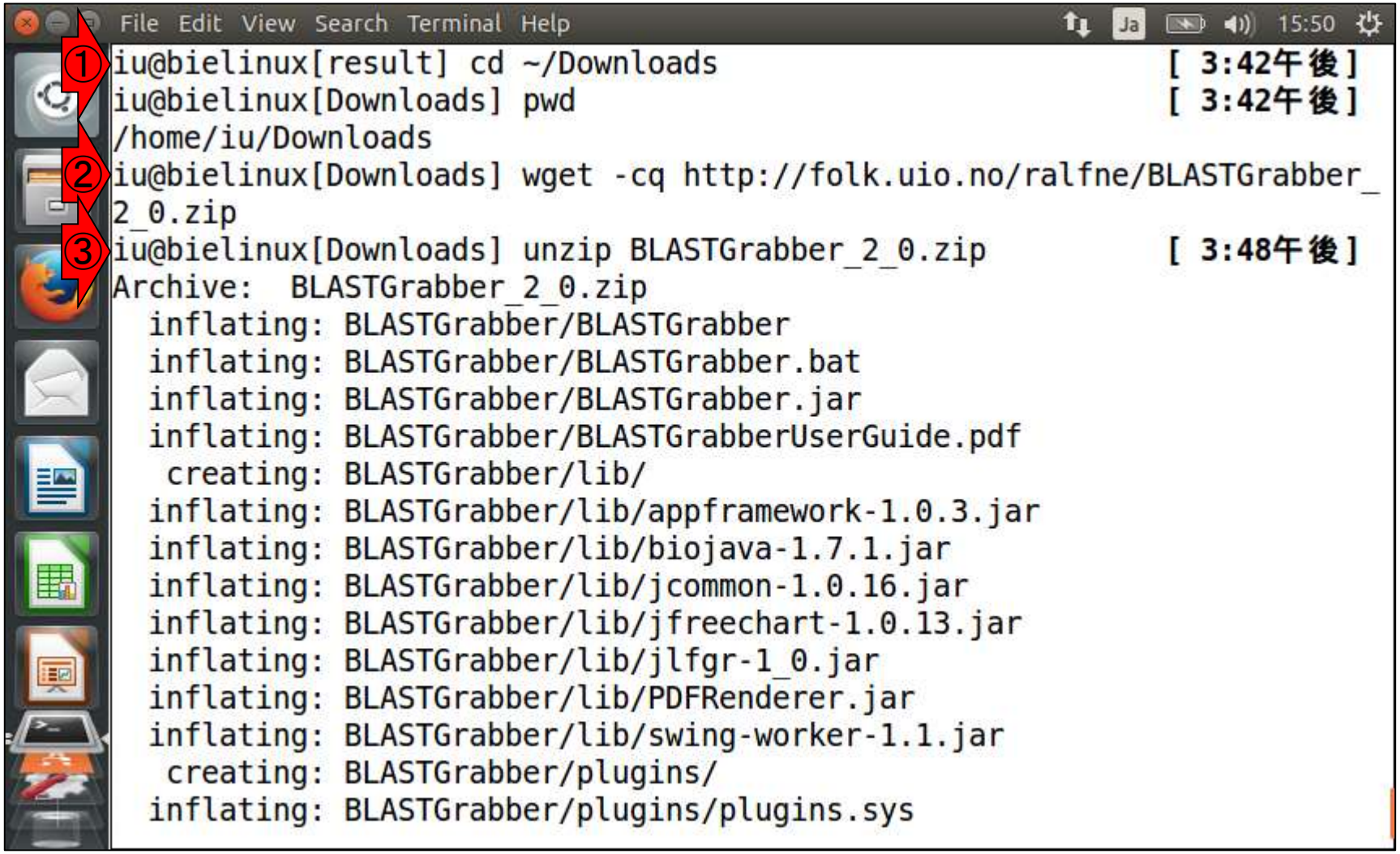
①BLASTGrabberのウェブページ。②BLASTGrabber 2.0からzip圧縮ファイルをダウンロード。右クリックでショートカットのコピー、でURL情報を取得してwgetするやり方が次のスライド



Neumann et al., *BMC Bioinformatics*, **15**: 128, 2014

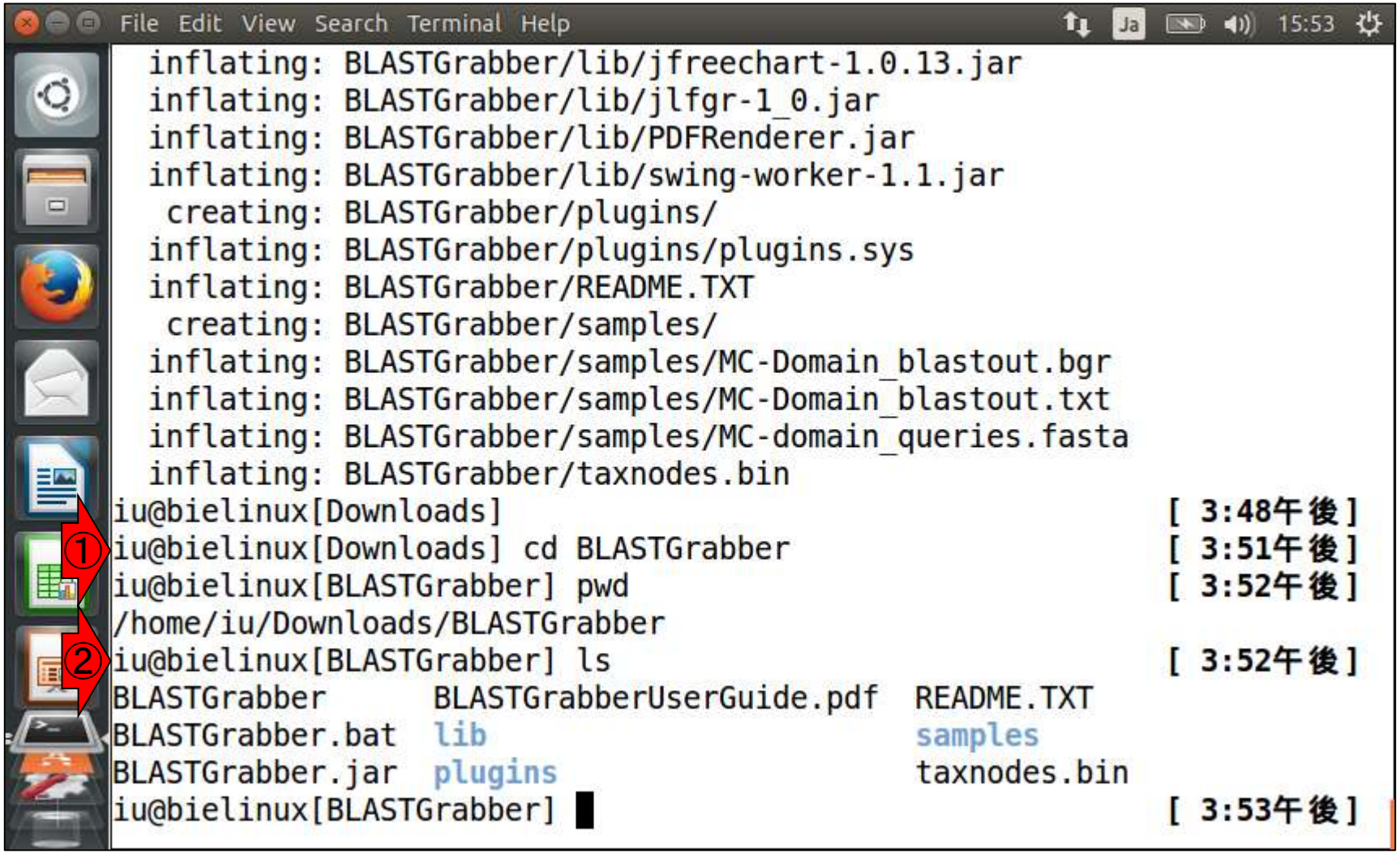
W7-1：BLASTGrabber

①~/Downloadsに移動し、②wget、③unzipで解凍



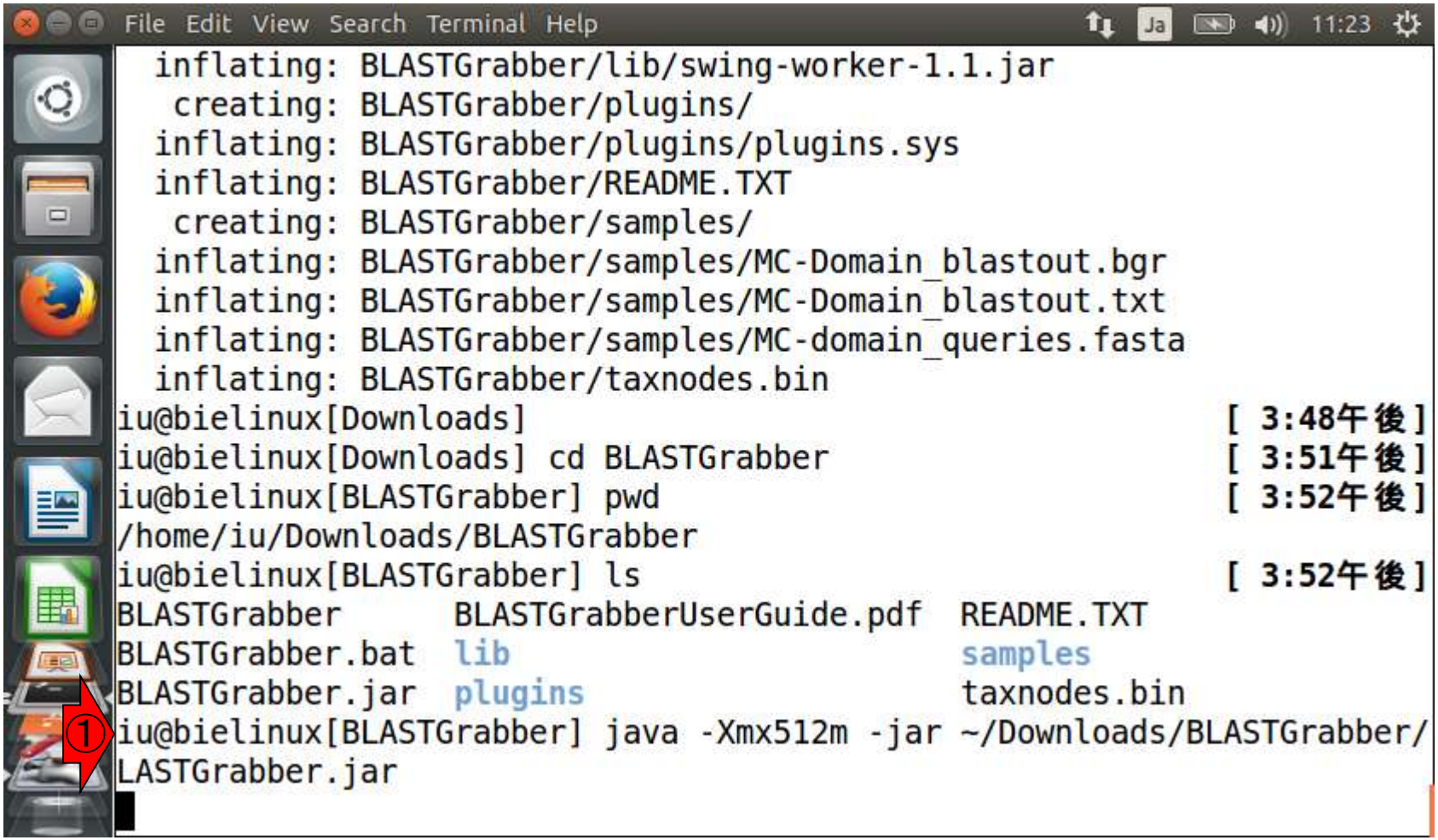
W7-1：BLASTGrabber

①解凍後に作成されたBLASTGrabberディレクトリに移動し、③ls



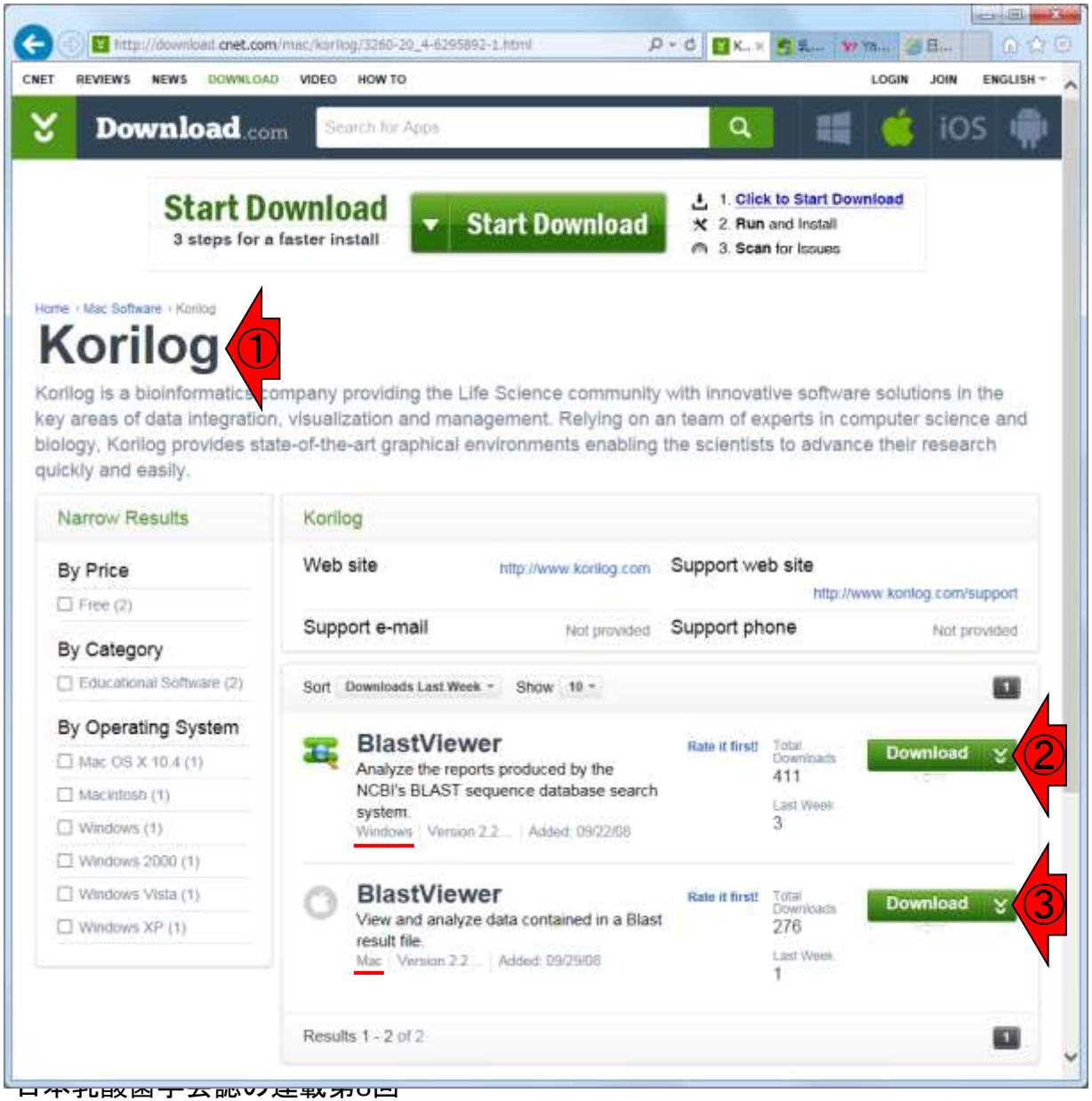
W7-1：BLASTGrabber

①BLASTGrabberの起動はこんな感じ。いろいろ試してみたが、直感的に使いづらく断念



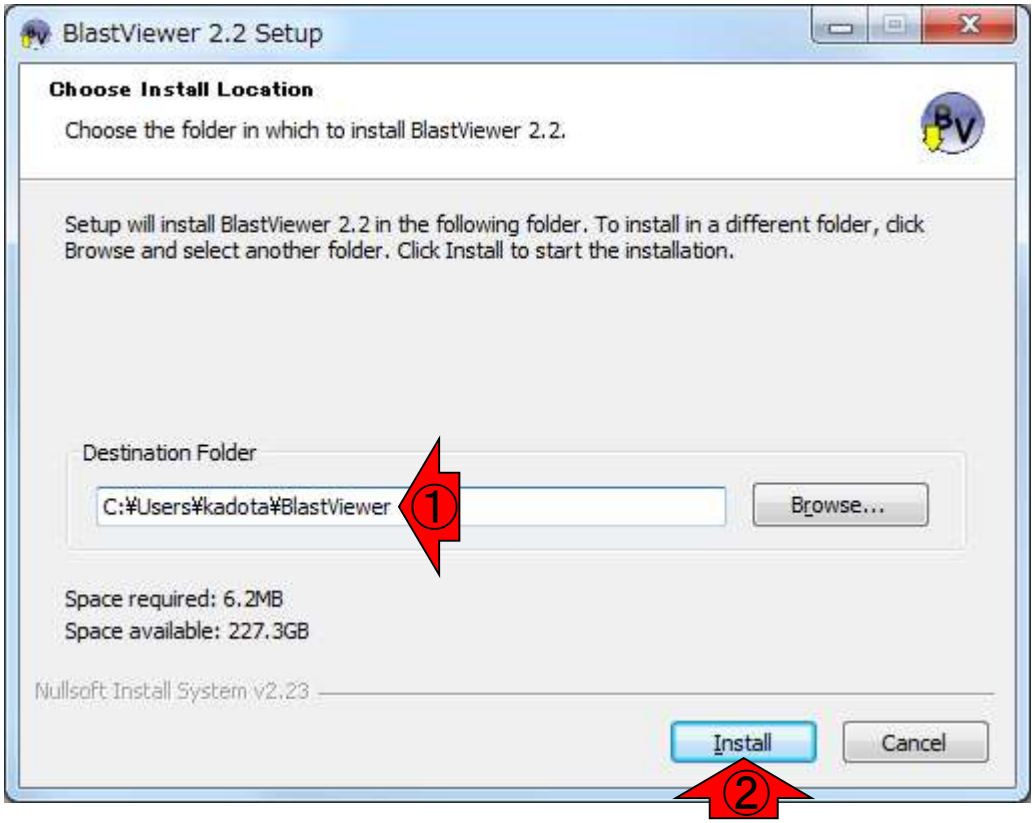
W7-2：BlastViewer

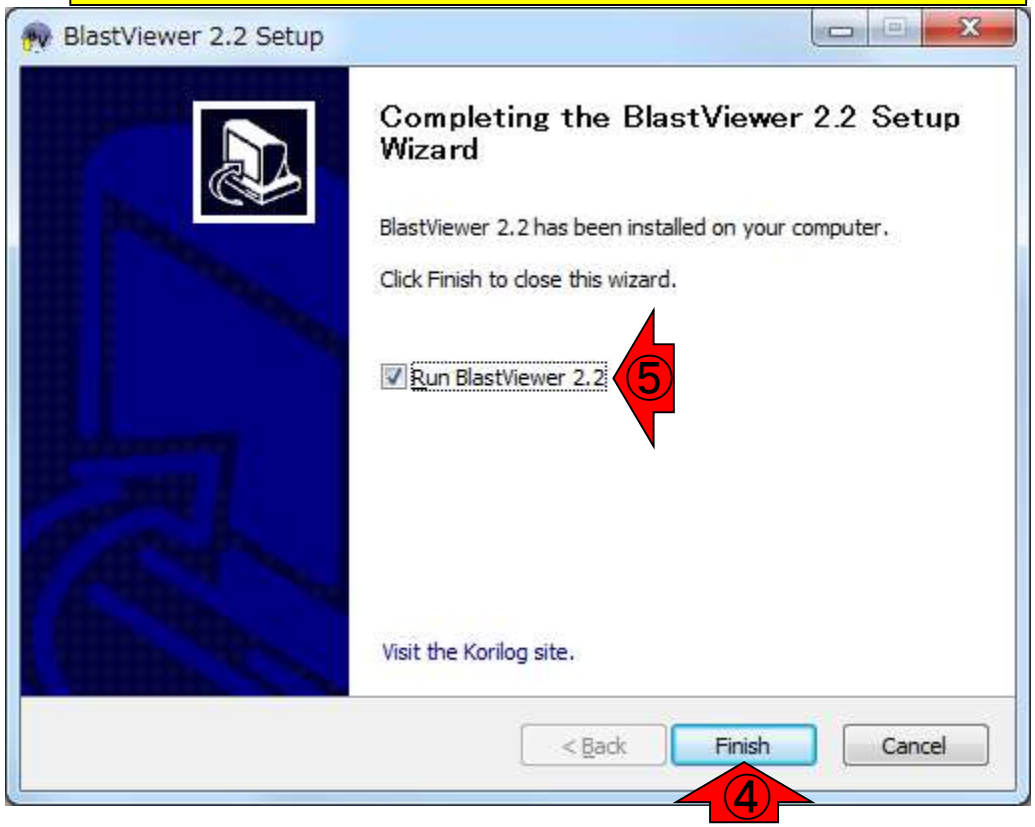
①Korilogというところが提供しているBlastViewerのウェブページ。②Windows版と③Macintosh版。次のスライド以降はWindows版で説明



W7-2：BlastViewer

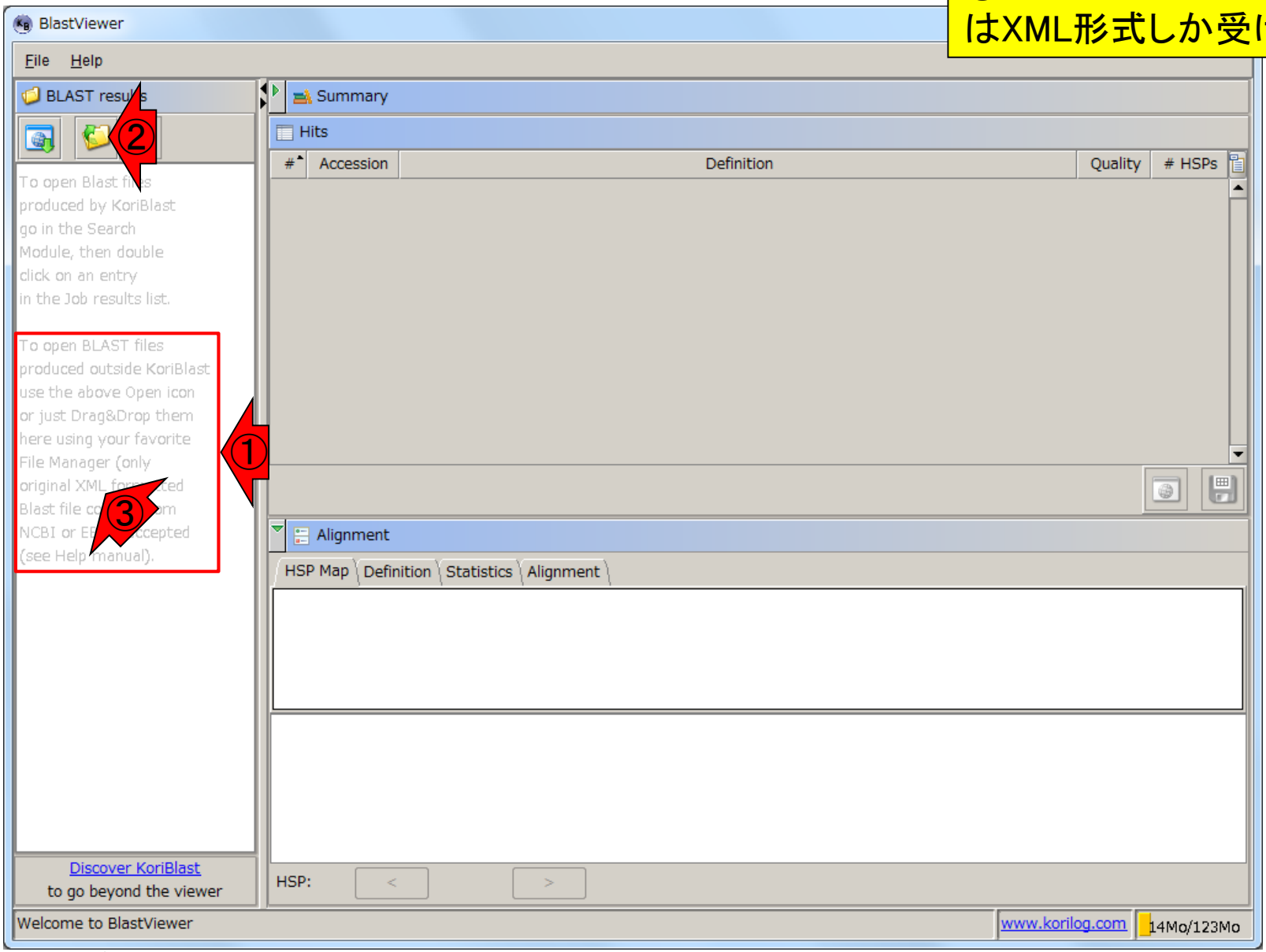
Windows 7上でのインストール画面。①ここに作成されるようだ。②Install。③デスクトップにBlastViewerのアイコンが作成されるはず。④Finish。デフォルトでは⑤にチェックが入っているのでBlastViewer (ver. 2.2)が起動する





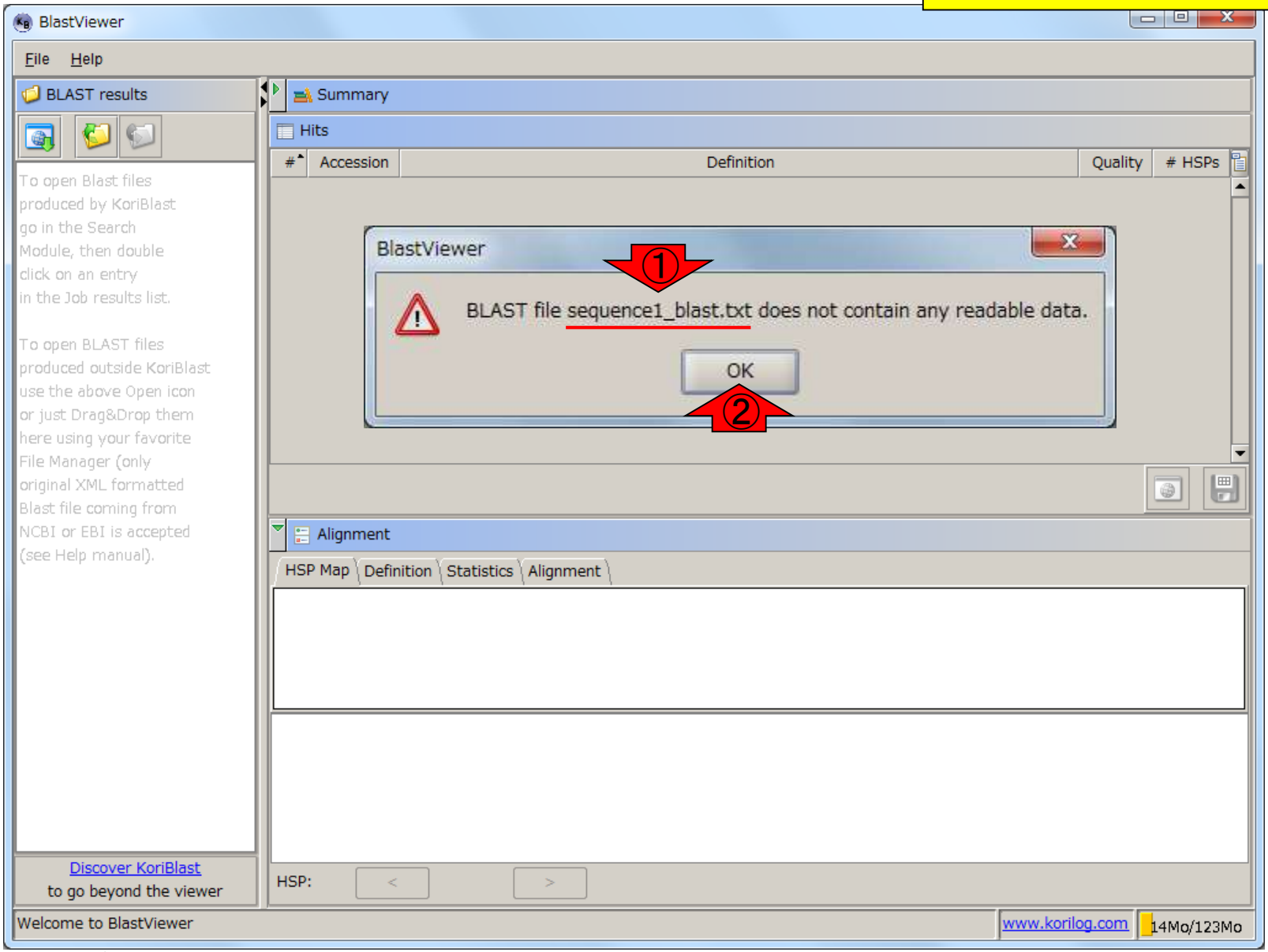
W7-2：BlastViewer

①を見ることで、Blast結果ファイルを開く際には、②を押せばよいことがわかる。③をよく見るとBLAST出力結果ファイルはXML形式しか受け付けていないようだ



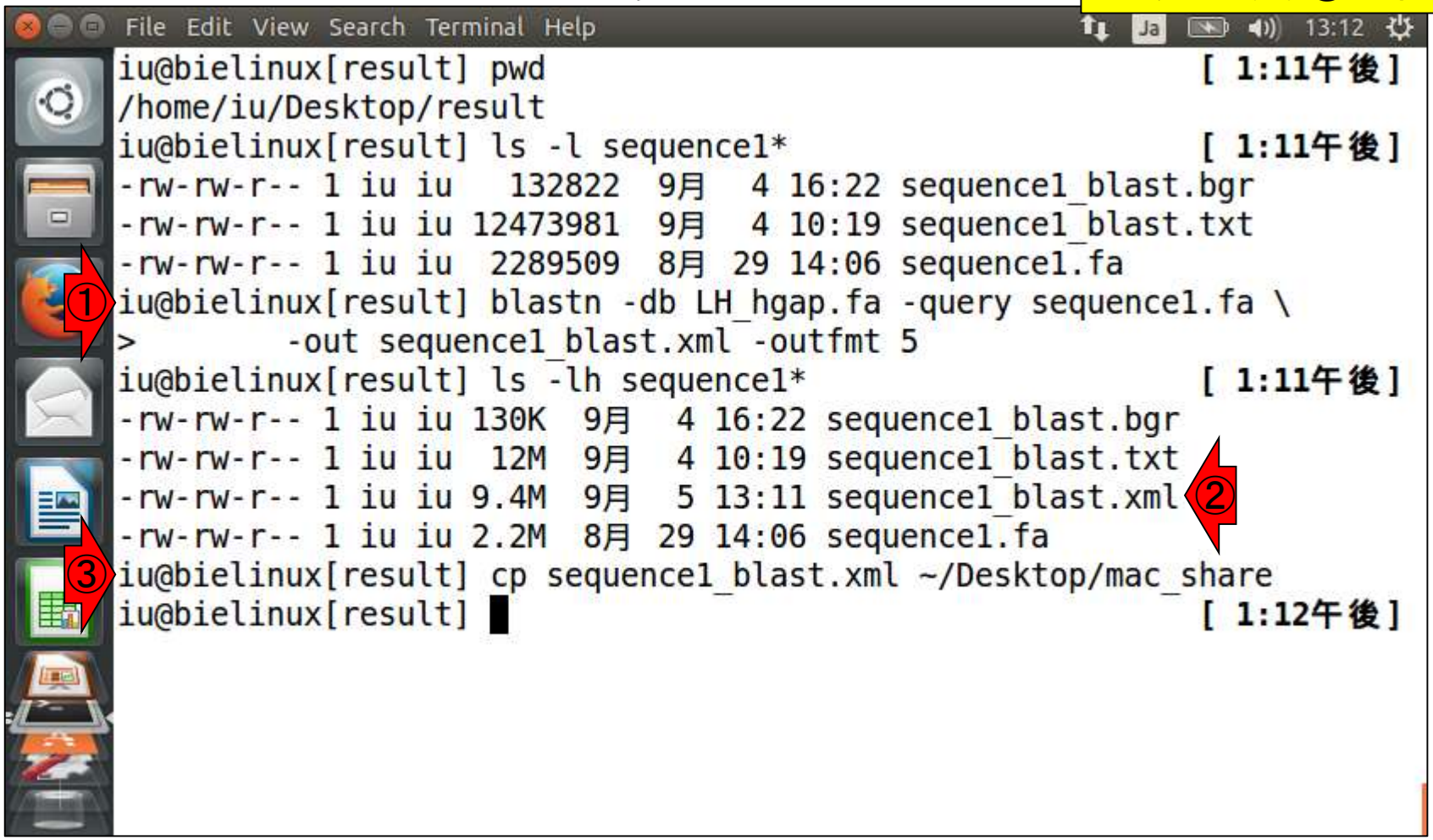
W7-2：BlastViewer

①ためしにW6-2で作成したテキスト形式のBlast結果ファイル(sequence1_blast.txt)を読み込もうとしたらダメでした。②OK



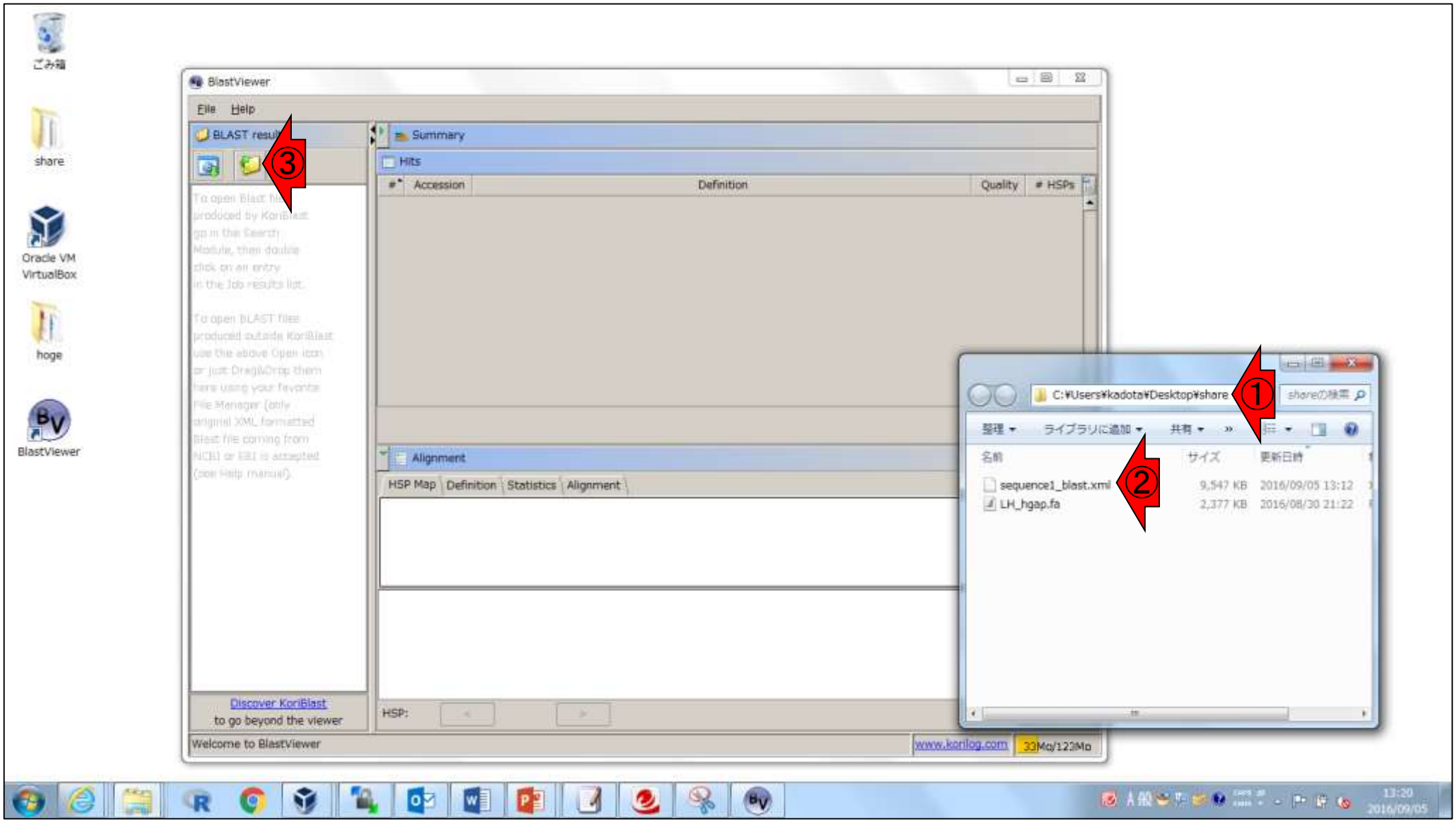
W7-3：Blast再実行

①XML形式のBlast結果ファイル(sequence1_blast.xml)を作成。②確かにあります。③共有フォルダにコピー



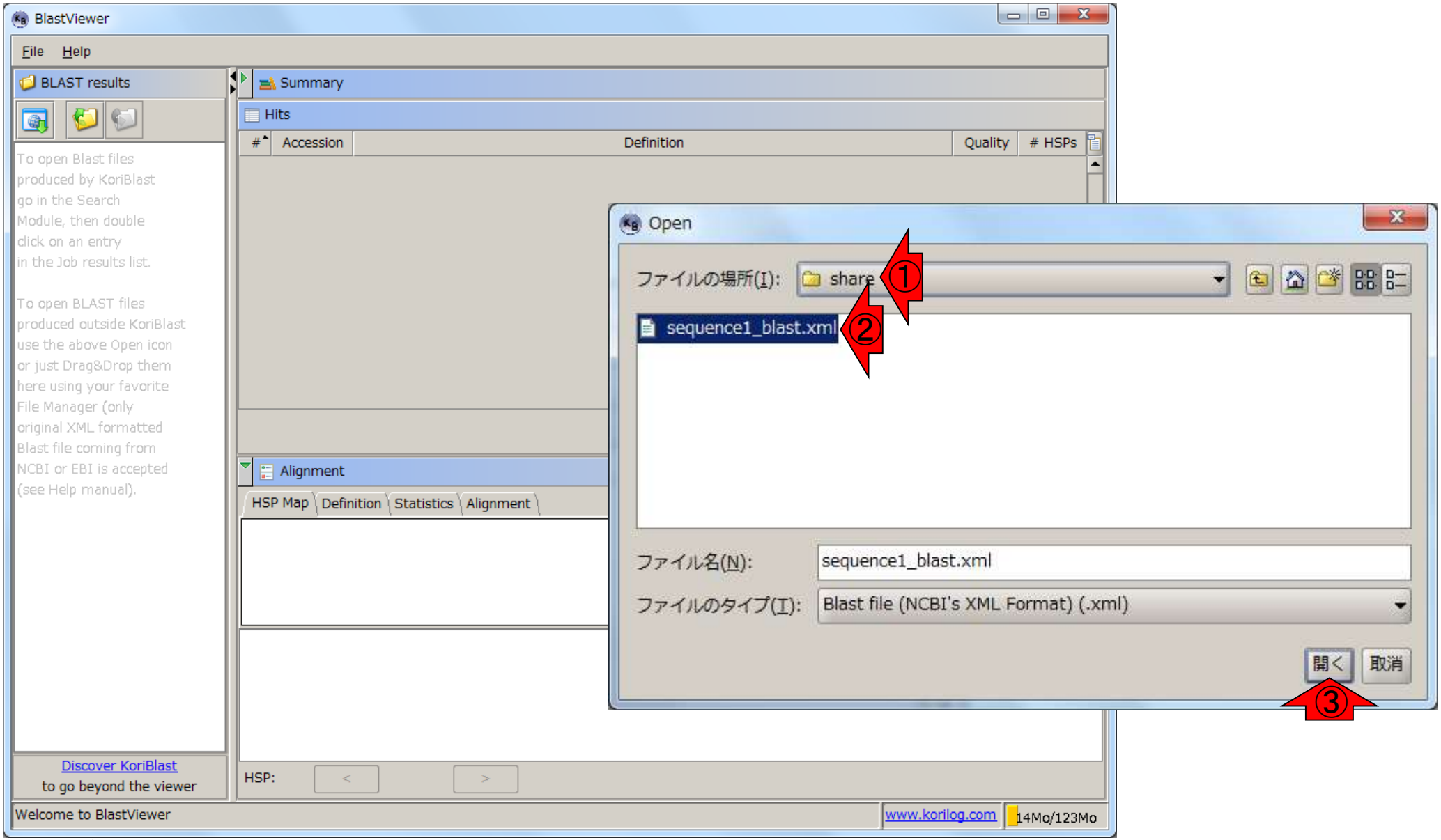
W7-4：BlastViewer

①共有フォルダ上の、②XML形式のBlast結果ファイル(sequence1_blast.xml)を、③開く



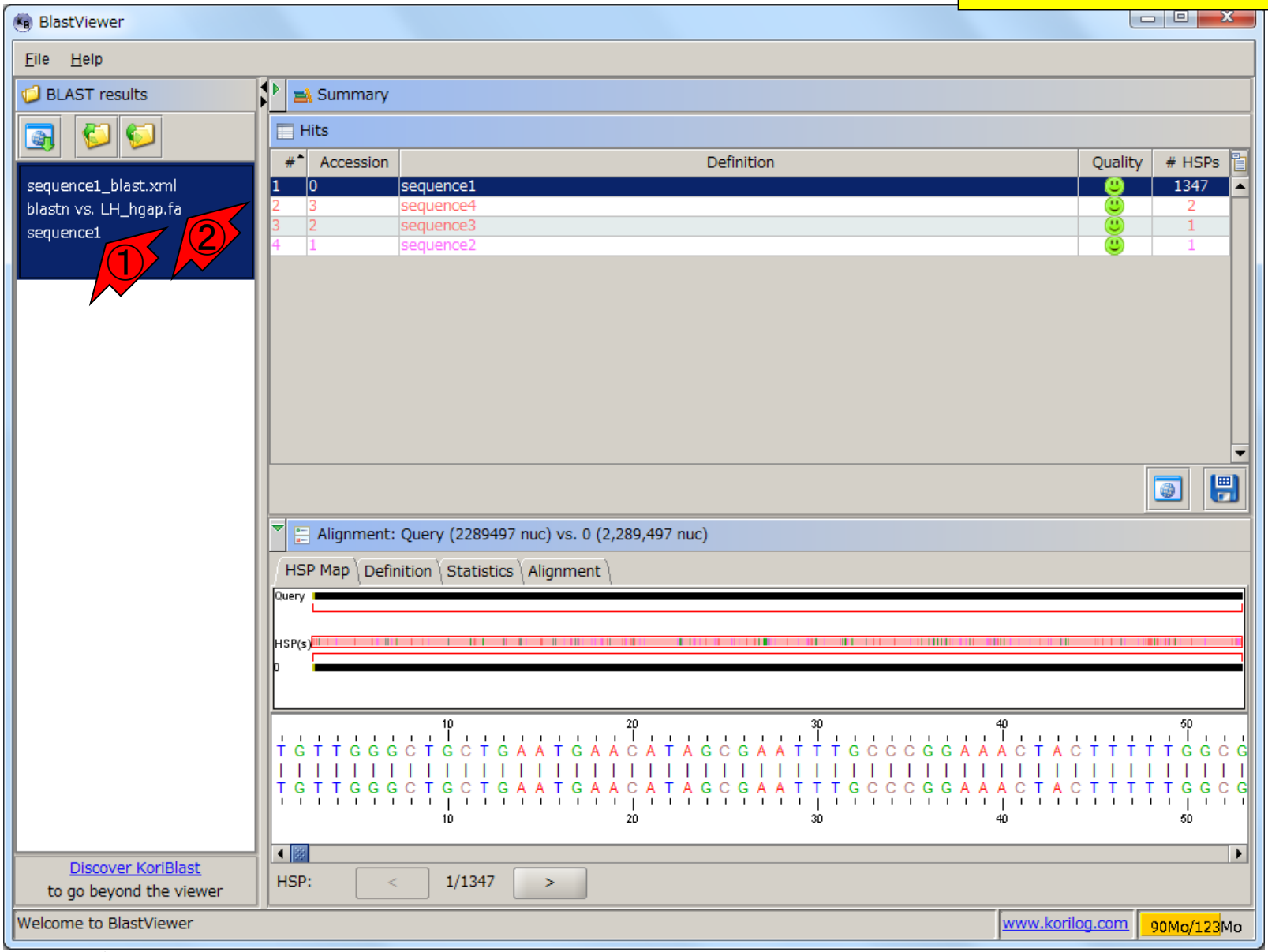
W7-4：BlastViewer

①共有フォルダ上の、②XML形式のBlast結果ファイル(sequence1_blast.xml)を、③開く



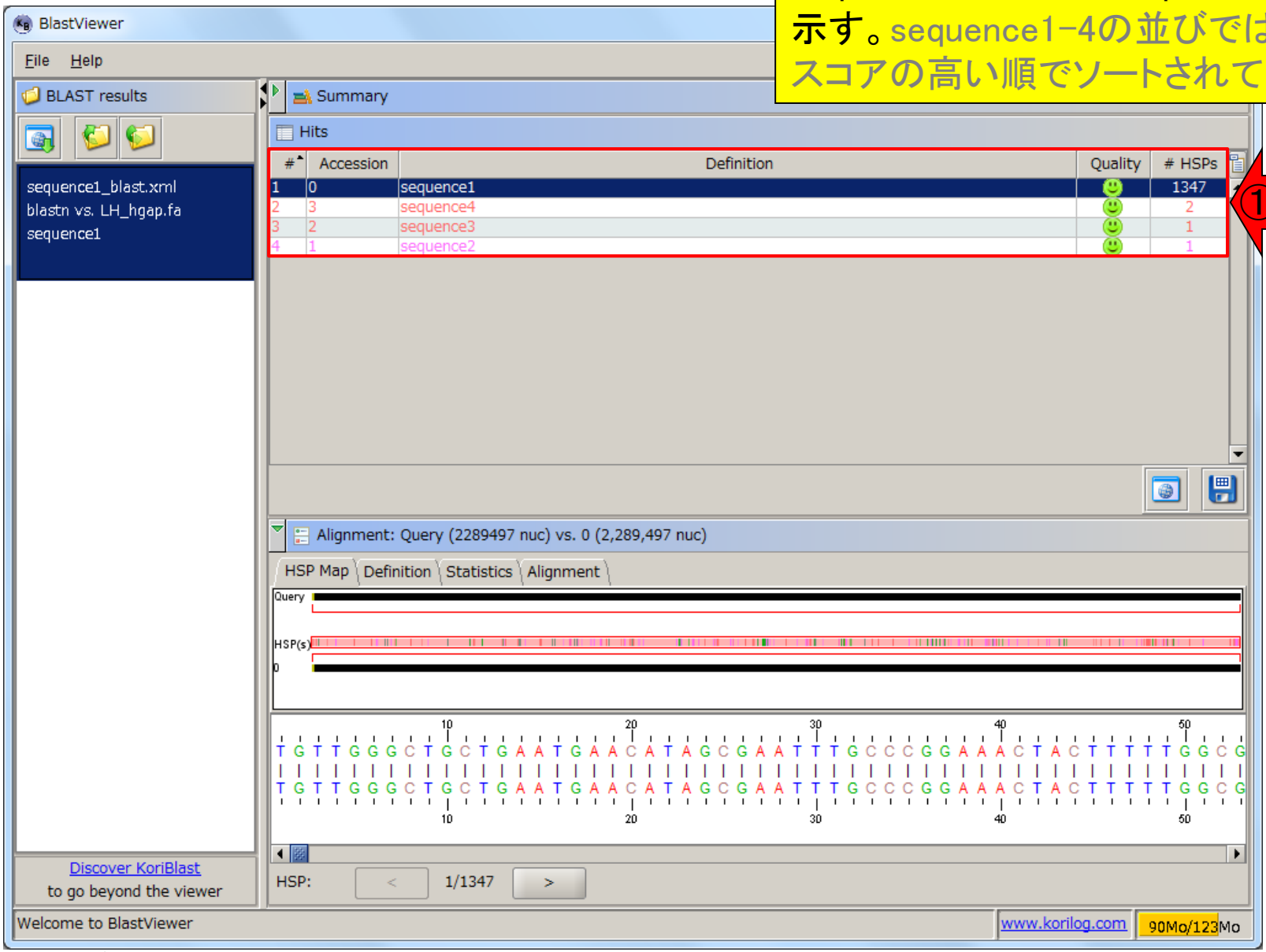
W7-4：BlastViewer

読み込み後の状態。①query側がsequence1.fa、②DB側がsequence1−4からなるLH_hgap.faだったことを思い出そう



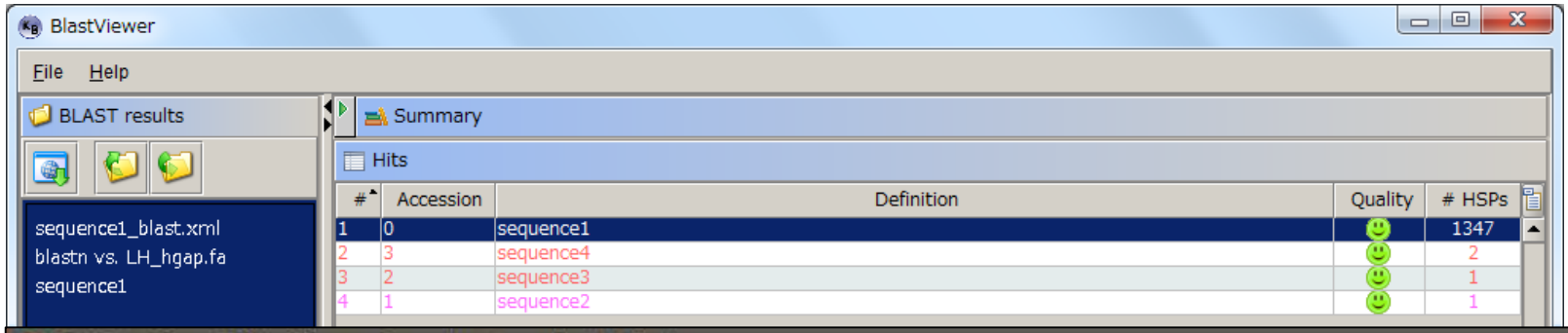
W7-5：解釈

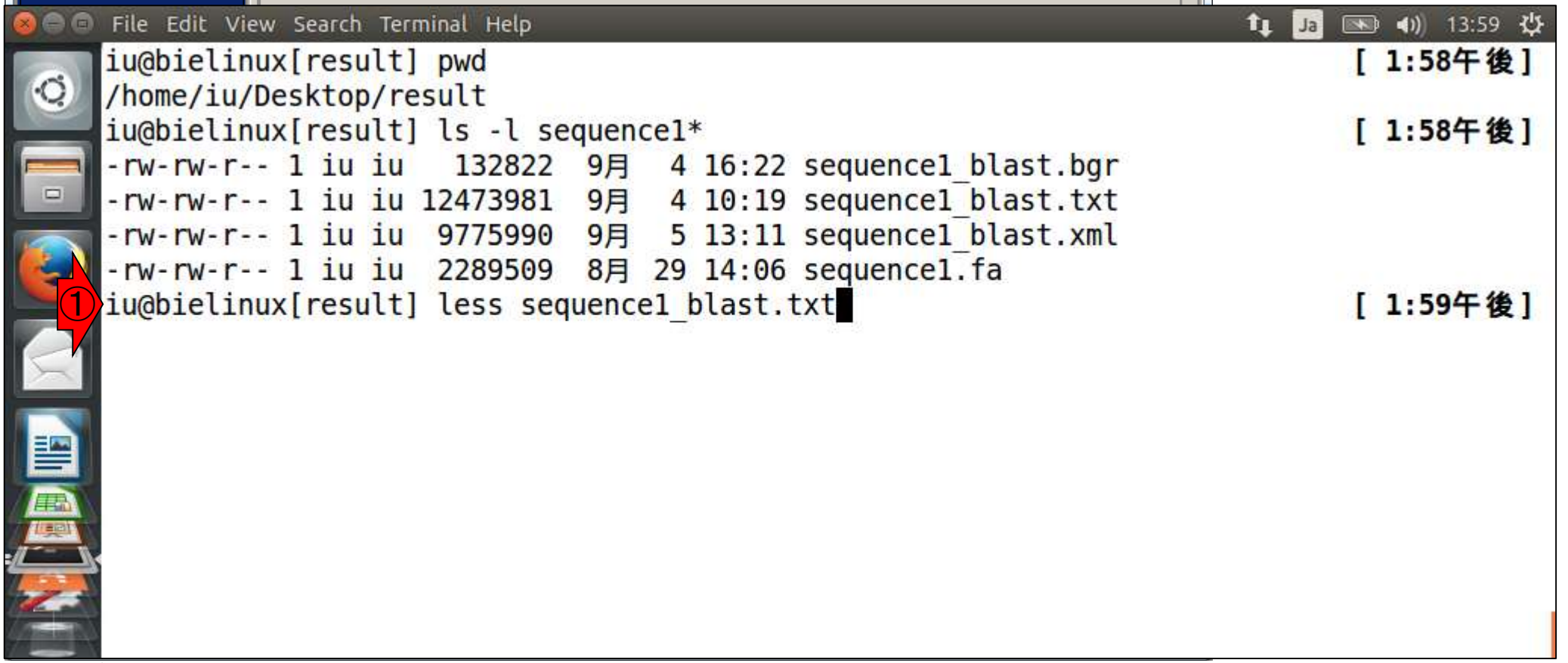
①は、DB側のヒット数（配列類似領域の数；HSPの数）が、sequence1中に1347個、sequence4中に2個、sequence3中に1個、sequence2中に1個あったことを示す。sequence1−4の並びではないので、HSP数かスコアの高い順でソートされているのだろうと妄想



W7-5：解釈

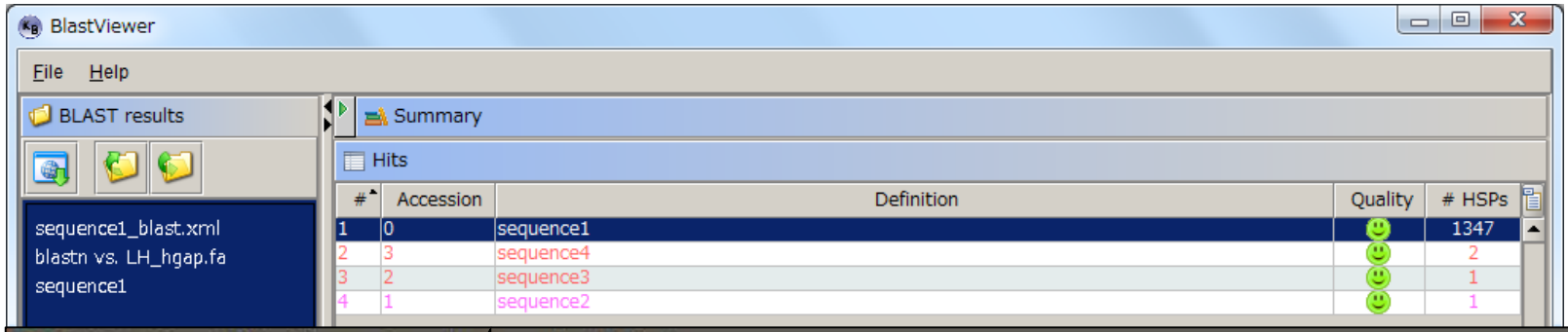
①テキスト形式のBLAST結果ファイル(sequence1_blast.txt)をlessで眺めて全体像を大まかに把握

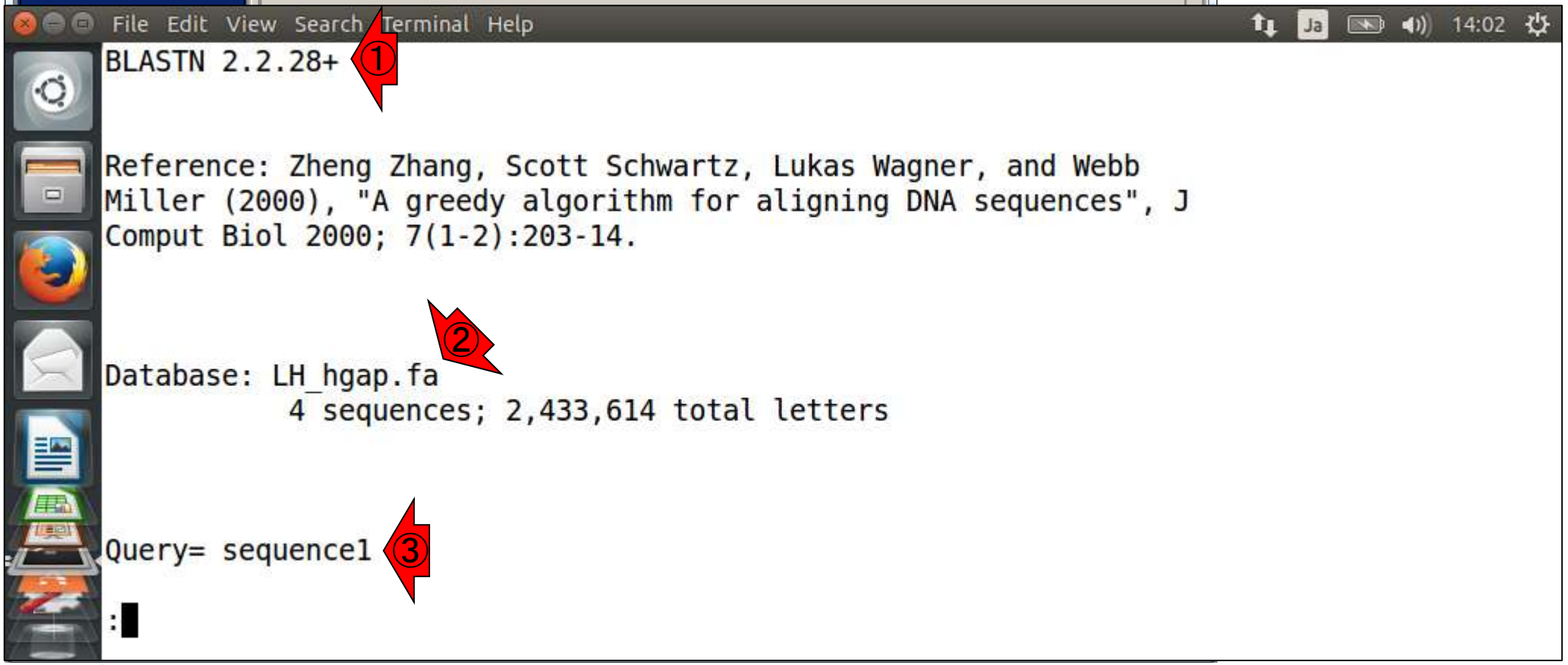




W7-5：解釈

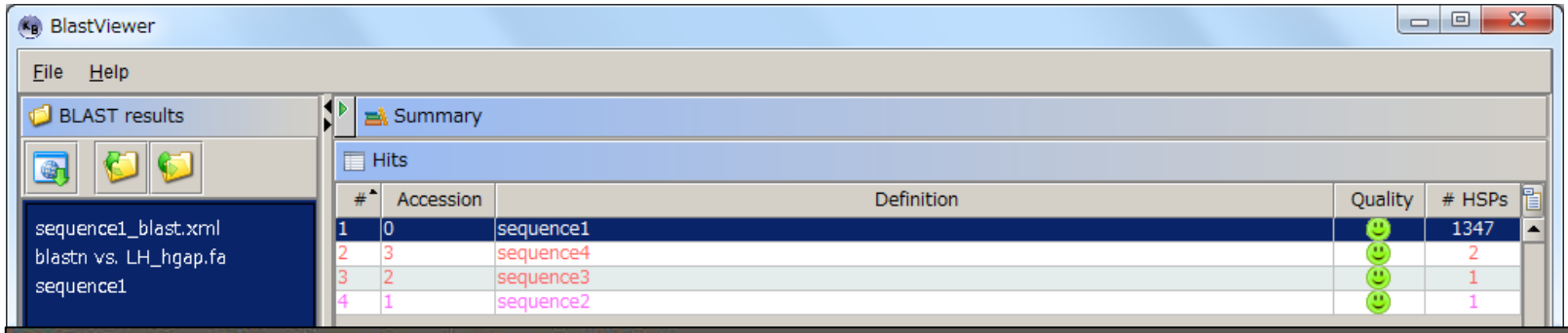
lessで開いた直後の状態。①用いたBLASTNのバージョン、②DB側、③query側の情報

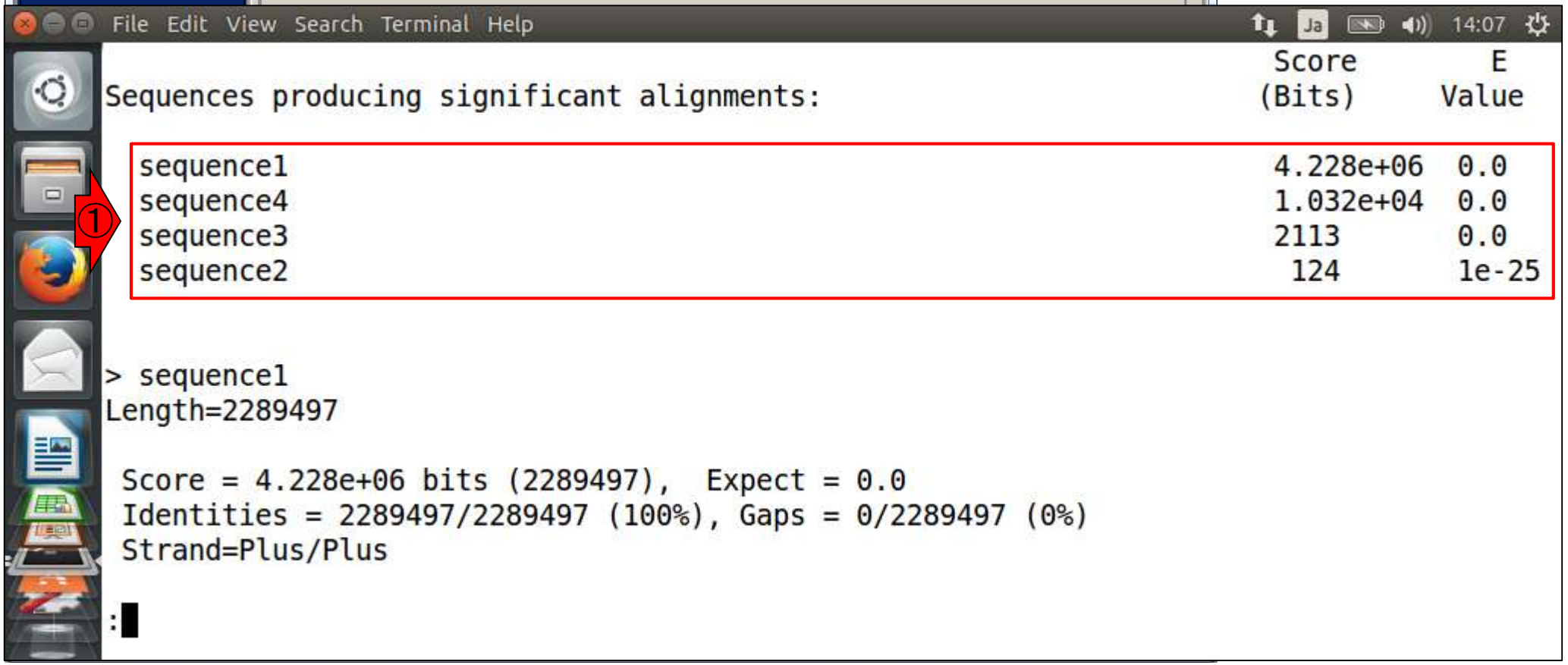




W7-5：解釈

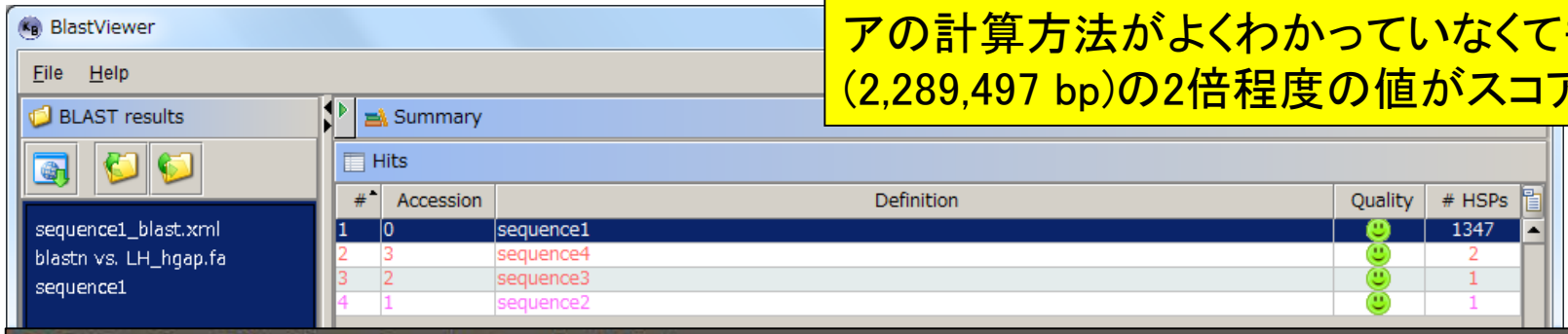
1ページ分ほど下に移動。①確かにBlastViewerと同じ並び(sequence1, 4, 3, 2)になっている

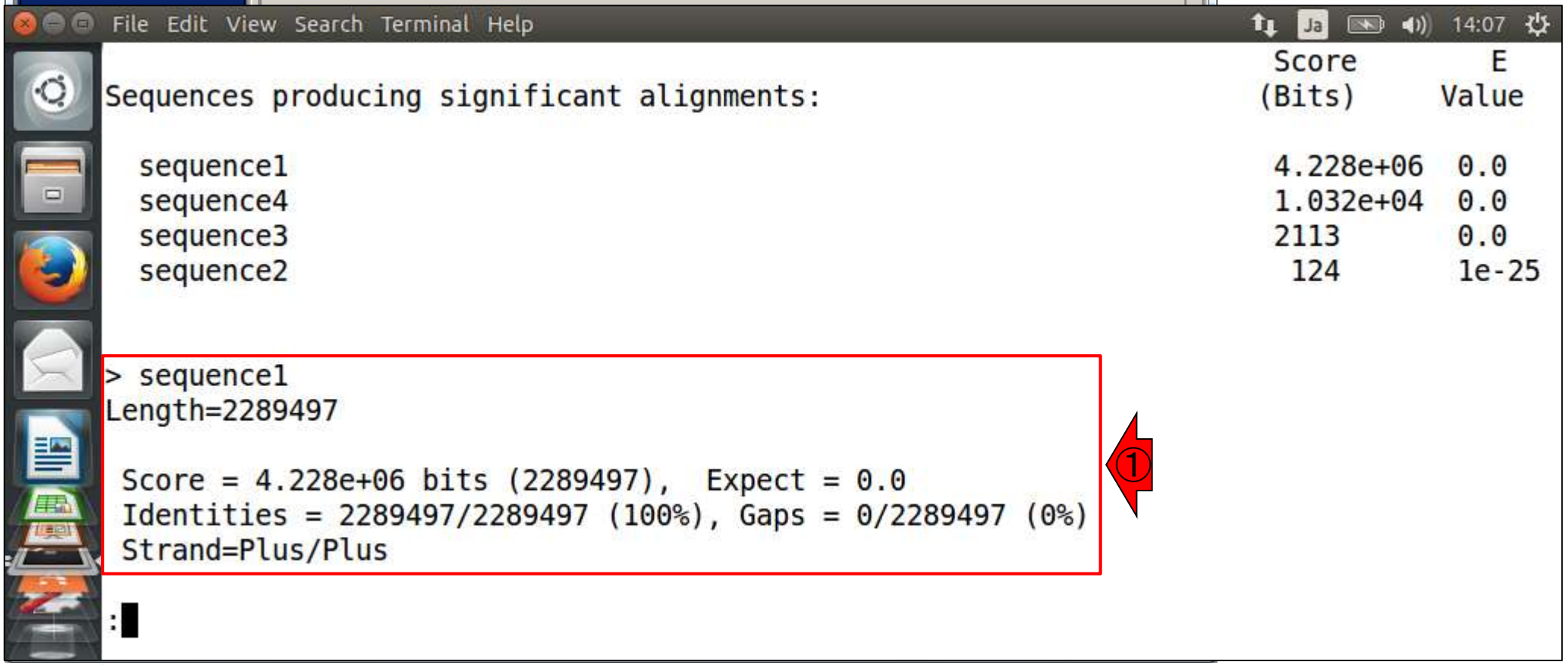




W7-5：解釈

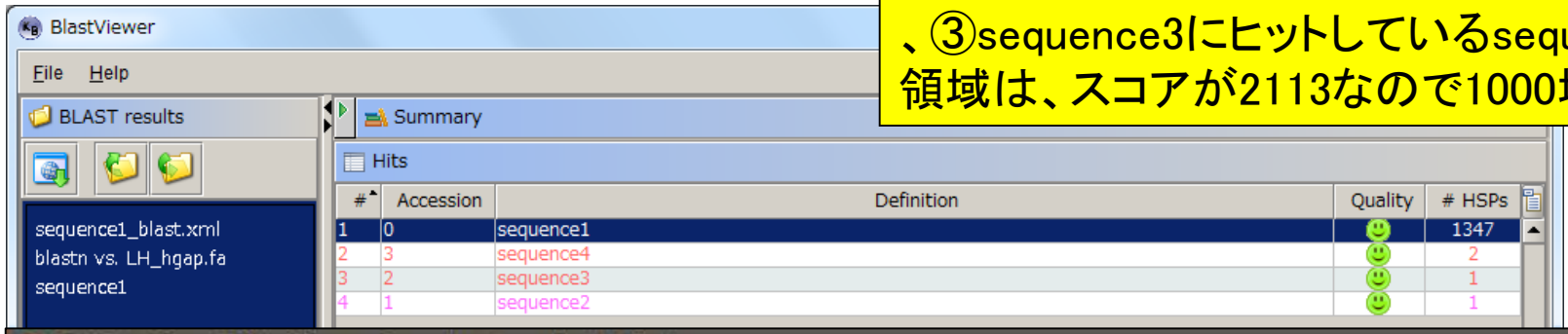
①これがトップヒットの基本情報。query（sequence1）とDB（sequence1-4）の関係が分かっていれば、これがsequence1 vs. sequence1の100%一致の結果であることがわかる。スコアの計算方法がよくわかっていなくても、大まかに配列長（2,289,497 bp）の2倍程度の値がスコアっぽいなと学習する

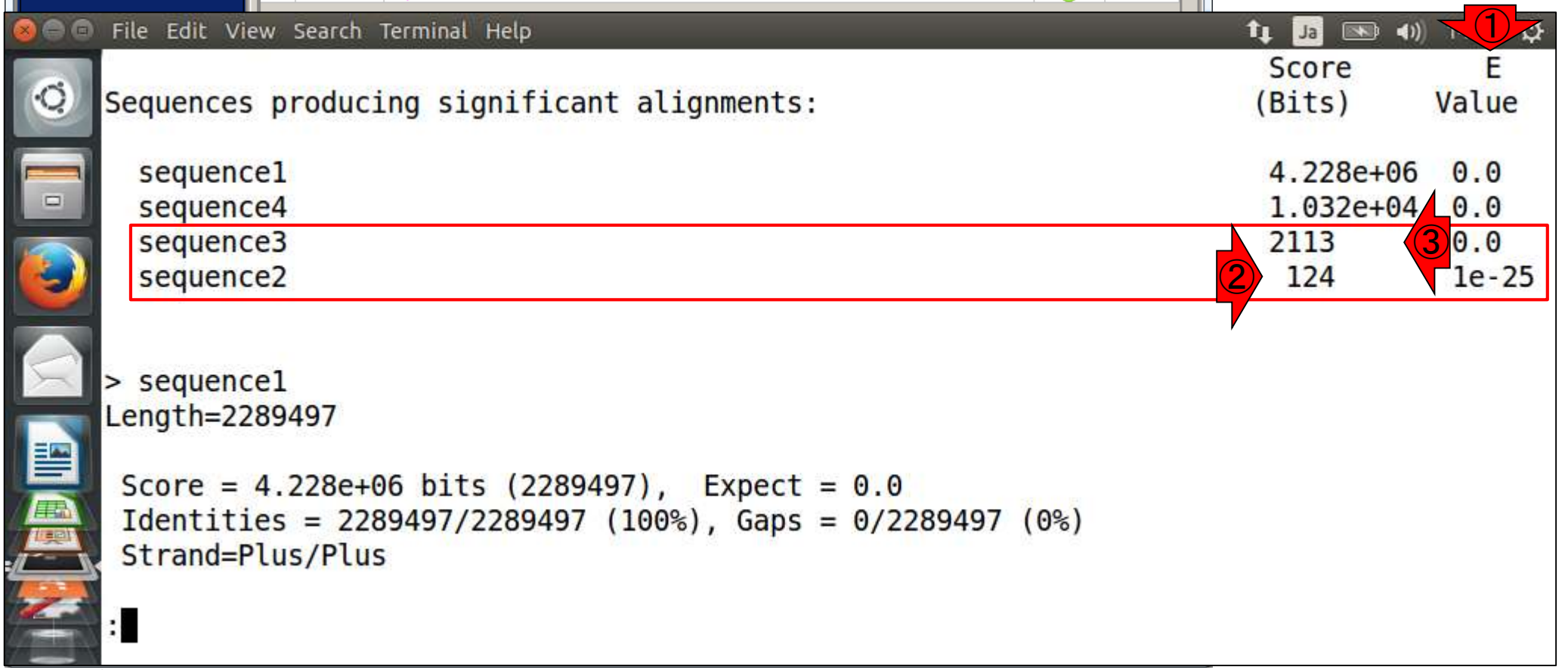




W7-5：解釈

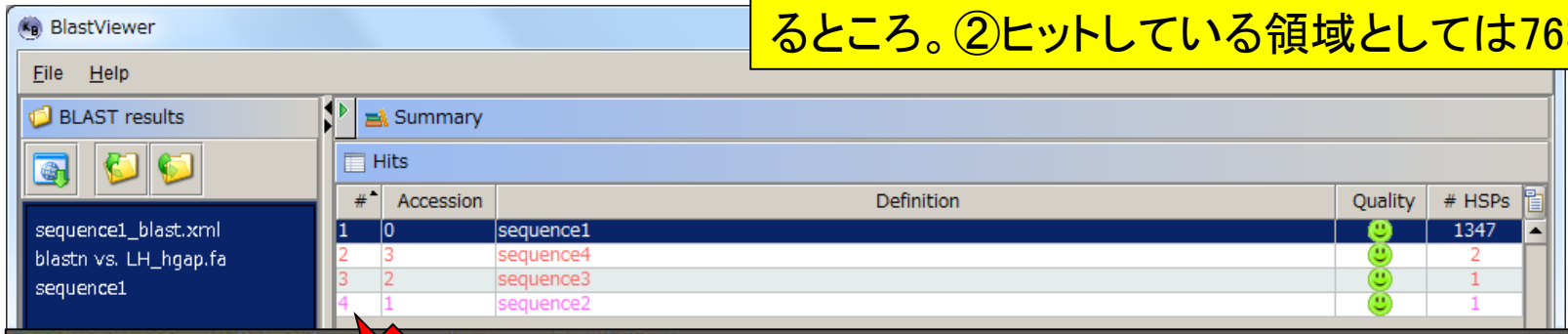
となると、いくら①E-valueがそこそこ低くても、②sequence2にヒットしているsequence1の断片配列の領域は、スコアが124なので60塩基程度だろうと予想。また、③sequence3にヒットしているsequence1の断片配列の領域は、スコアが2113なので1000塩基程度だろうと予想

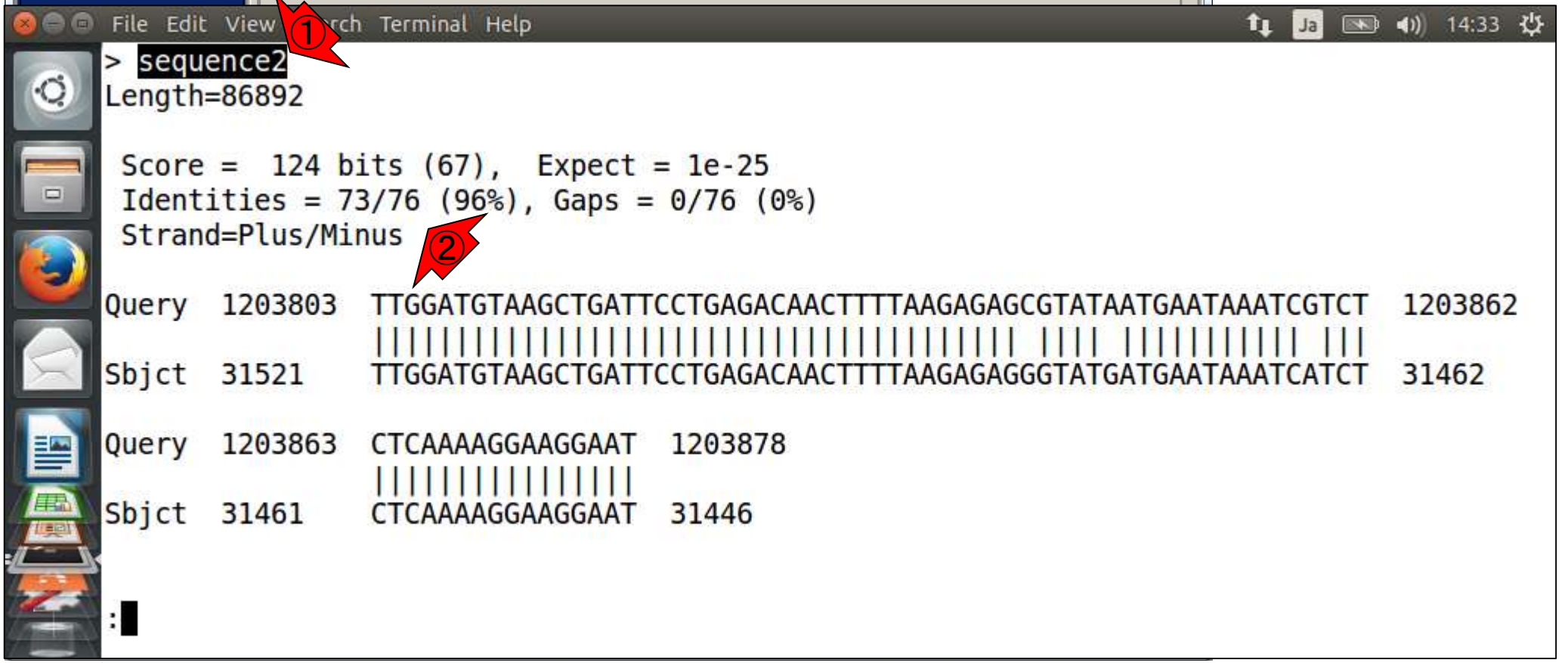




W7-5：解釈

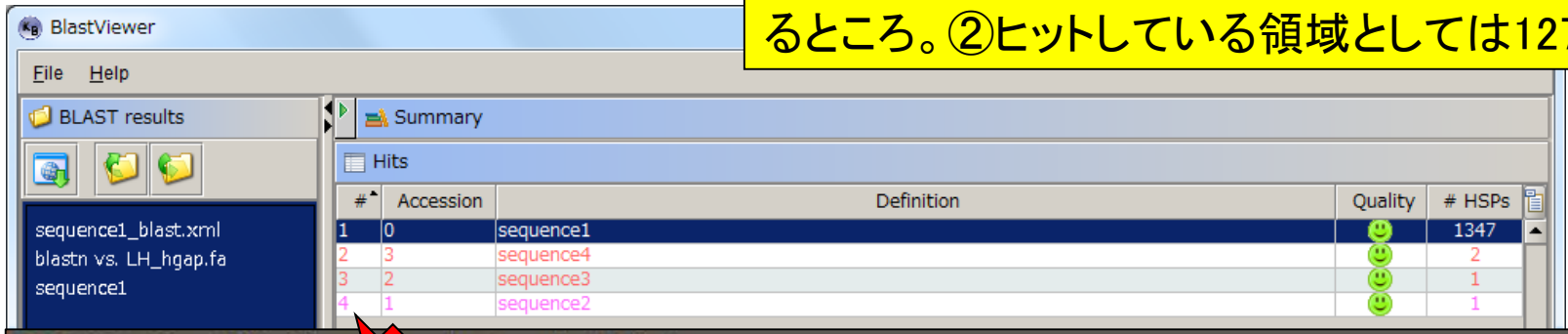
今は、テキスト形式のBLAST結果ファイル(sequence1_blast.txt)をlessで眺めている。①sequence2でキーワード検索（nで順方向に検索、Nで逆方向に検索）し、アラインメントを表示させているところ。②ヒットしている領域としては76塩基

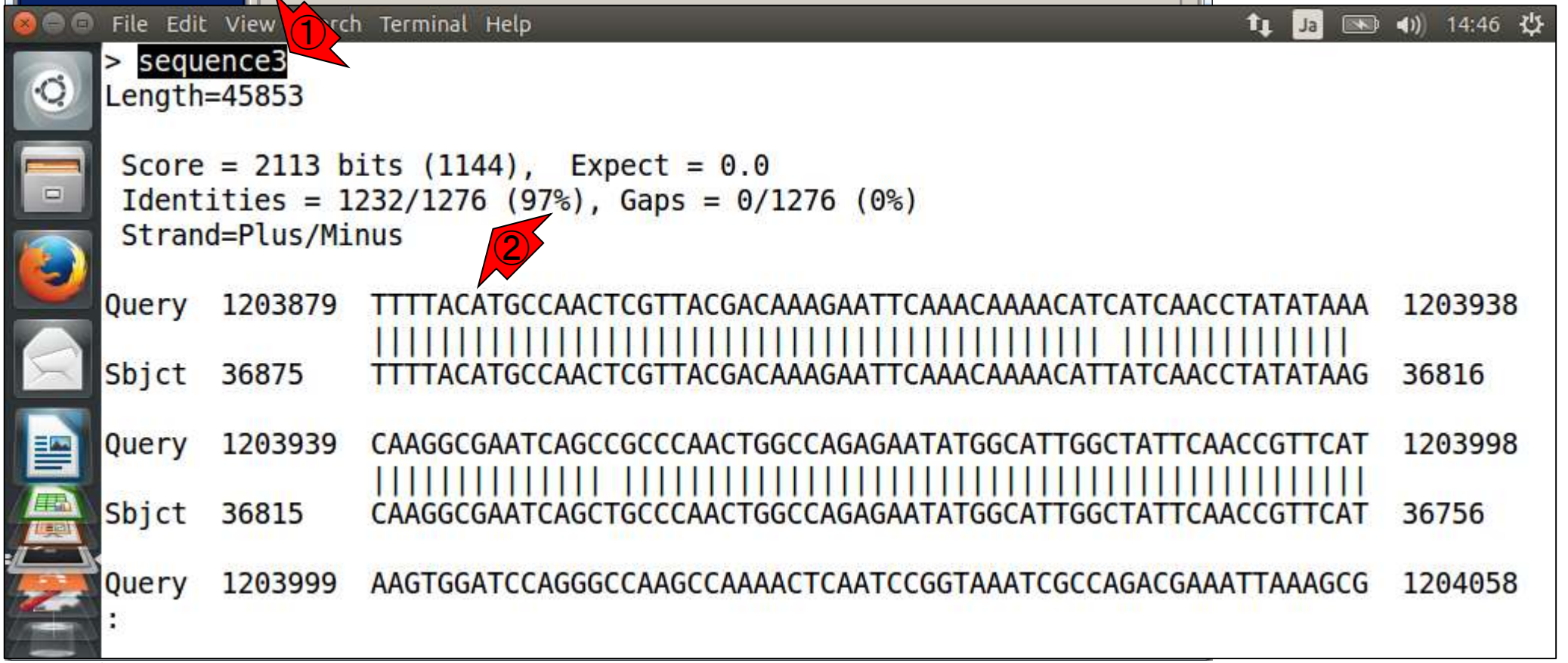




W7-5：解釈

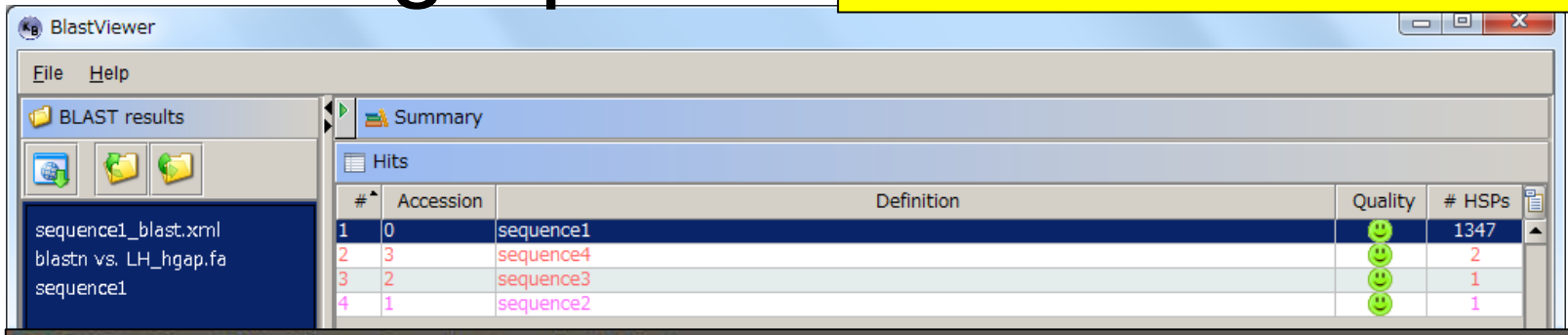
今は、テキスト形式のBLAST結果ファイル(sequence1_blast.txt)をlessで眺めている。①sequence3でキーワード検索（nで順方向に検索、Nで逆方向に検索）し、アラインメントを表示させているところ。②ヒットしている領域としては1276塩基

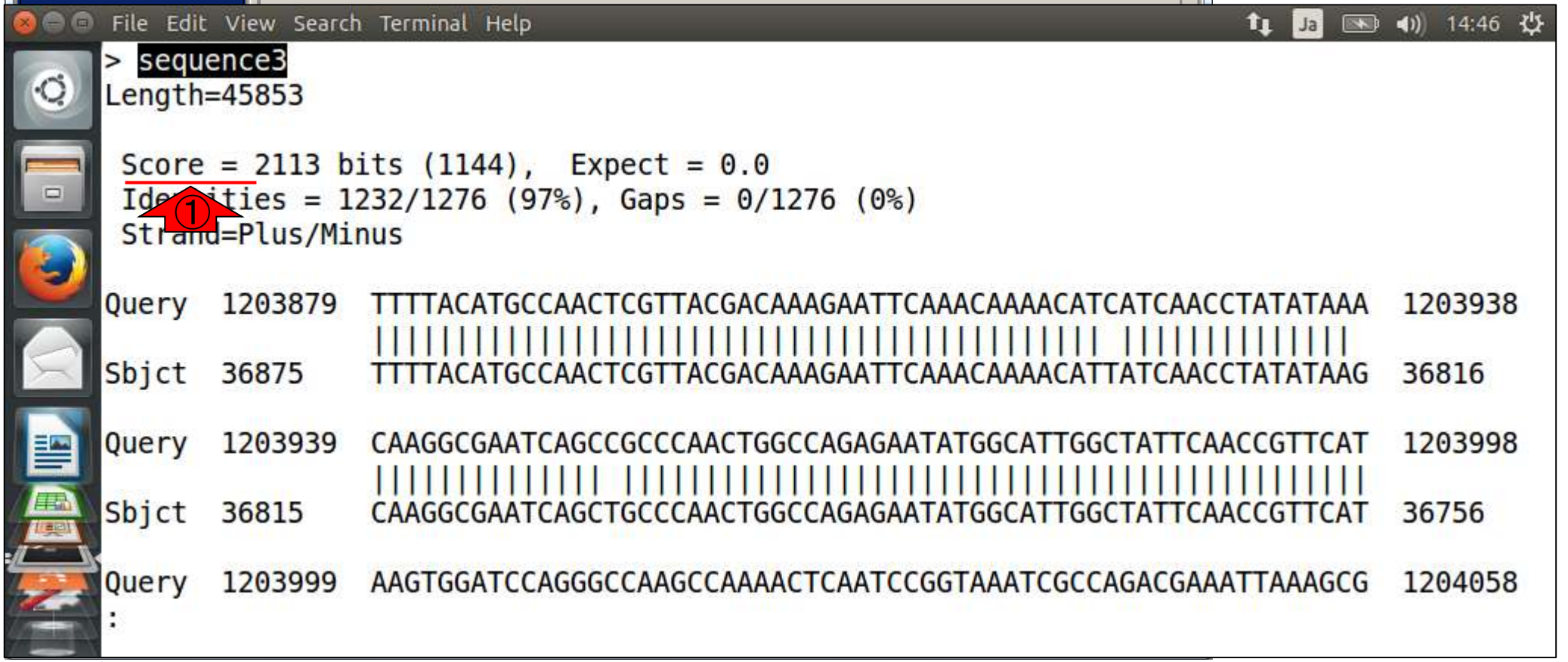




W7-6：grep

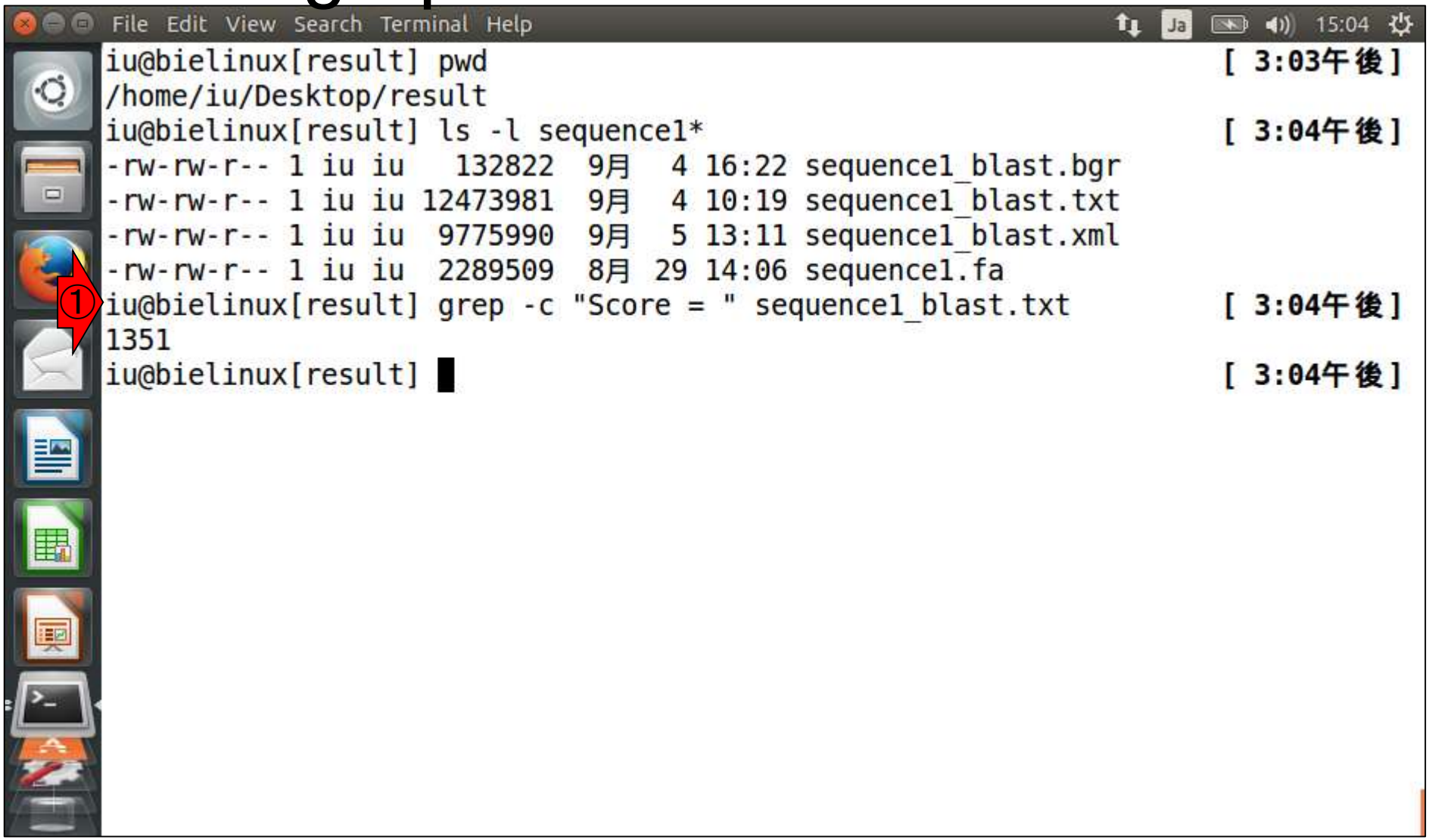
トータルのヒット数が(1347 + 2 + 1 + 1) = 1351個だったということは、①「Score = 」を含む行数も1351個あるのだろうと予想。grepで確認すべく、qを押してlessから一旦抜ける





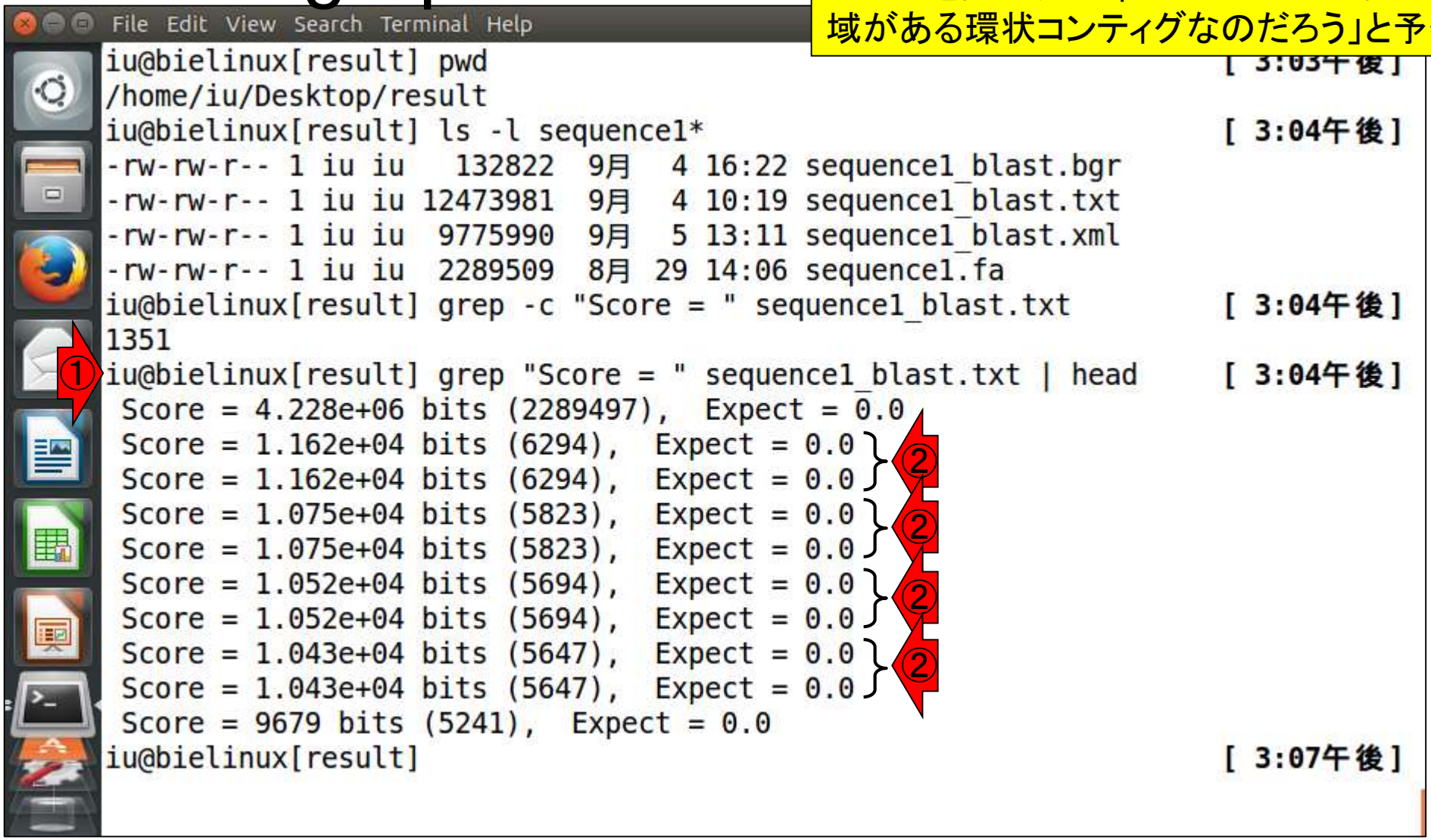
①「Score = 」を含む行数をgrepで調査。確かに1351個ある

W7-6：grep



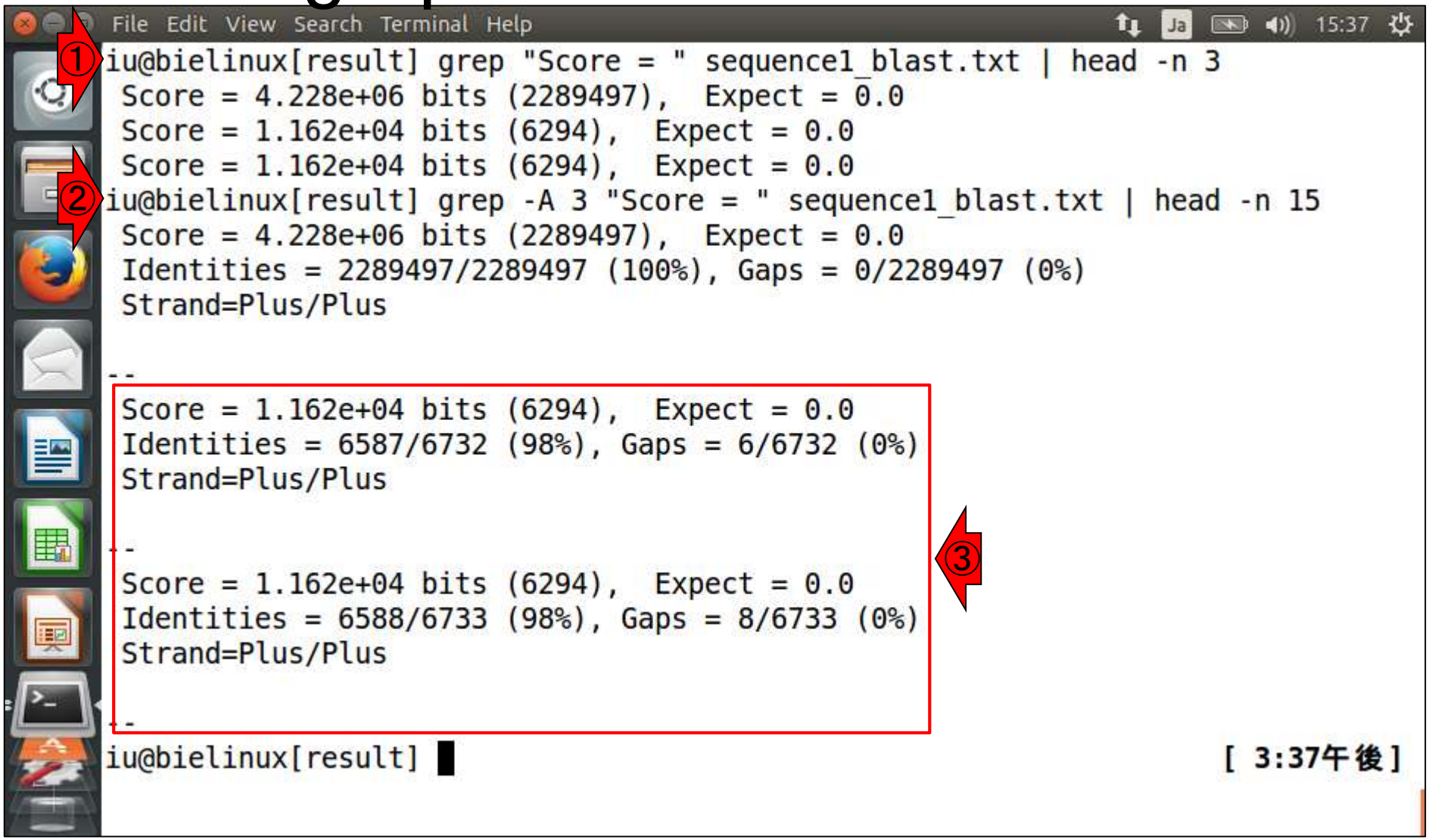
W7-6：grep

①「Score = 」を含む最初の10行分を表示。②のように同じスコアのものが2個づつ表示されている状態を見て、「sequence1も両末端に重複領域がある環状コンティグなのだろう」と予想する



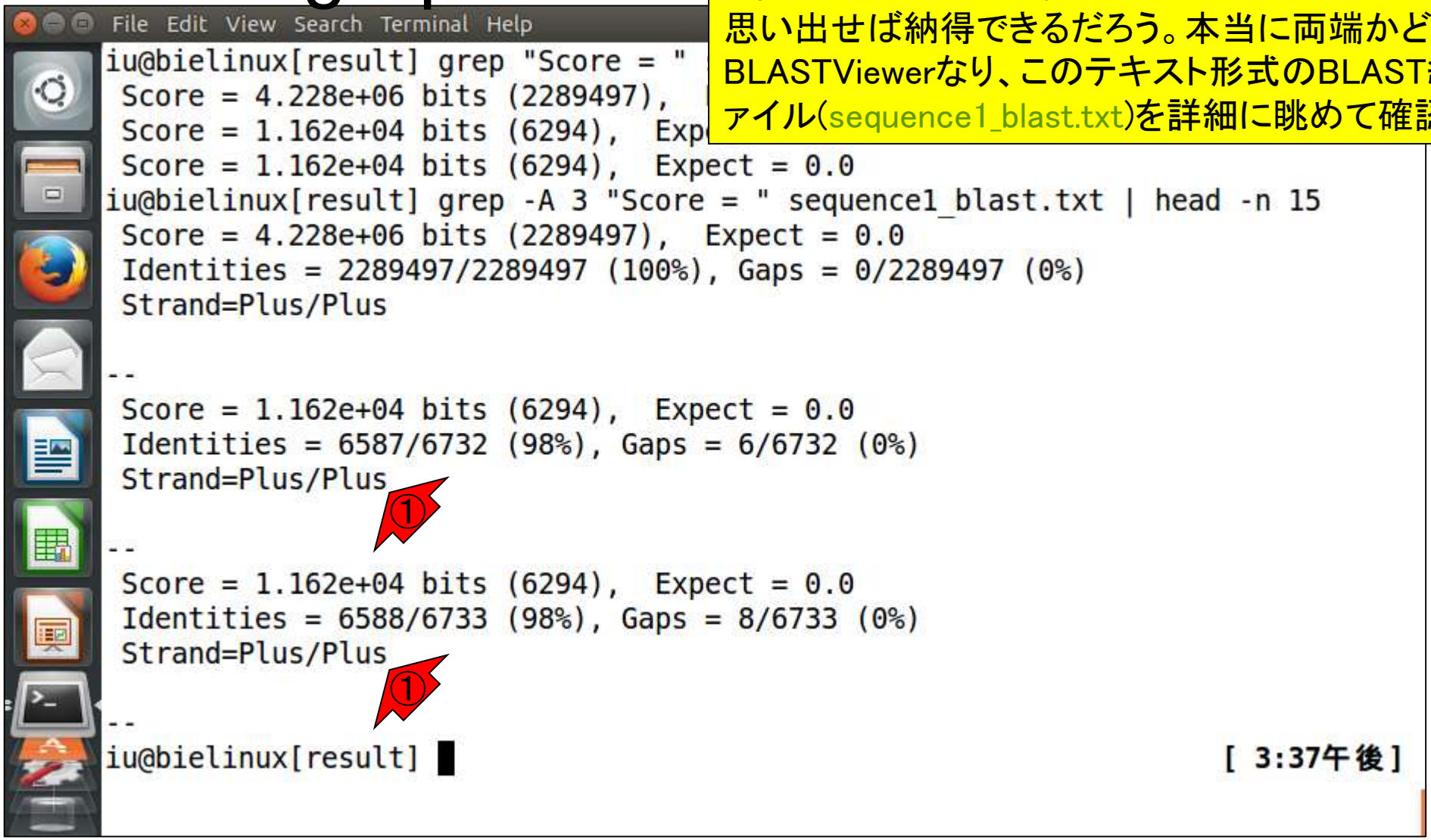
W7-6：grep

①「Score = 」を含む最初の3行分を表示。②grep -Aオプションで一致した行を含め後ろの3行分を表示。③が今注目しているところ



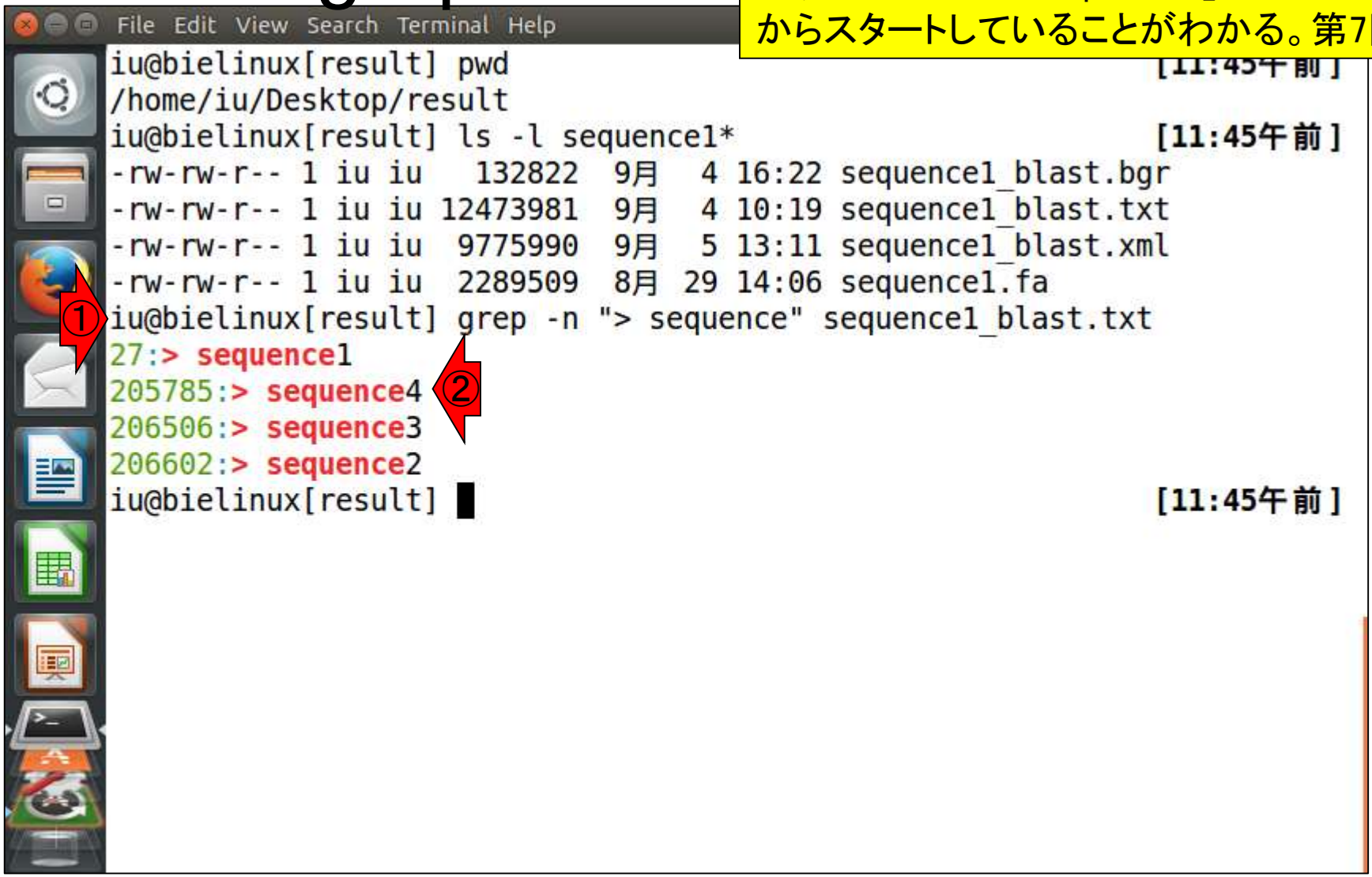
W7-6：grep -A

①Plus/Plusになっているので、おそらく両端に重複領域があるのだろう。sequence3同士のBLAST実行結果（第7回W15-6）も両端でPlus/Plusになっていたことを思い出せば納得できるだろう。本当に両端かどうかはBLASTViewerなり、このテキスト形式のBLAST結果ファイル(sequence1_blast.txt)を詳細に眺めて確認する



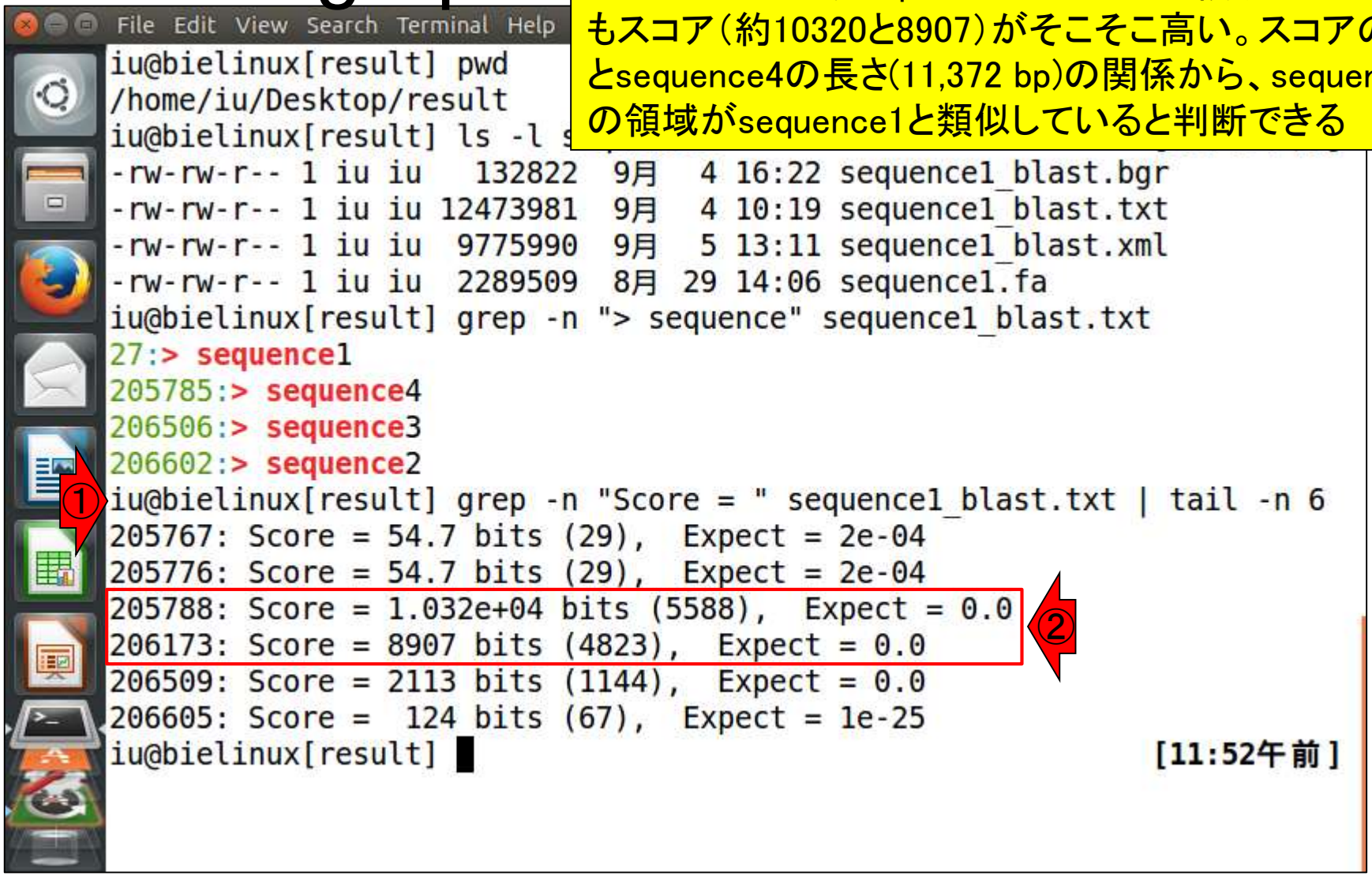
W7-6：grep -n

①grep -nで検索文字列（この場合”> sequence”）を含む行番号を表示。例えば②sequence4とヒットしたものは、入力ファイル(sequence1_blast.txt)の205785行目からスタートしていることがわかる。第7回W15-5



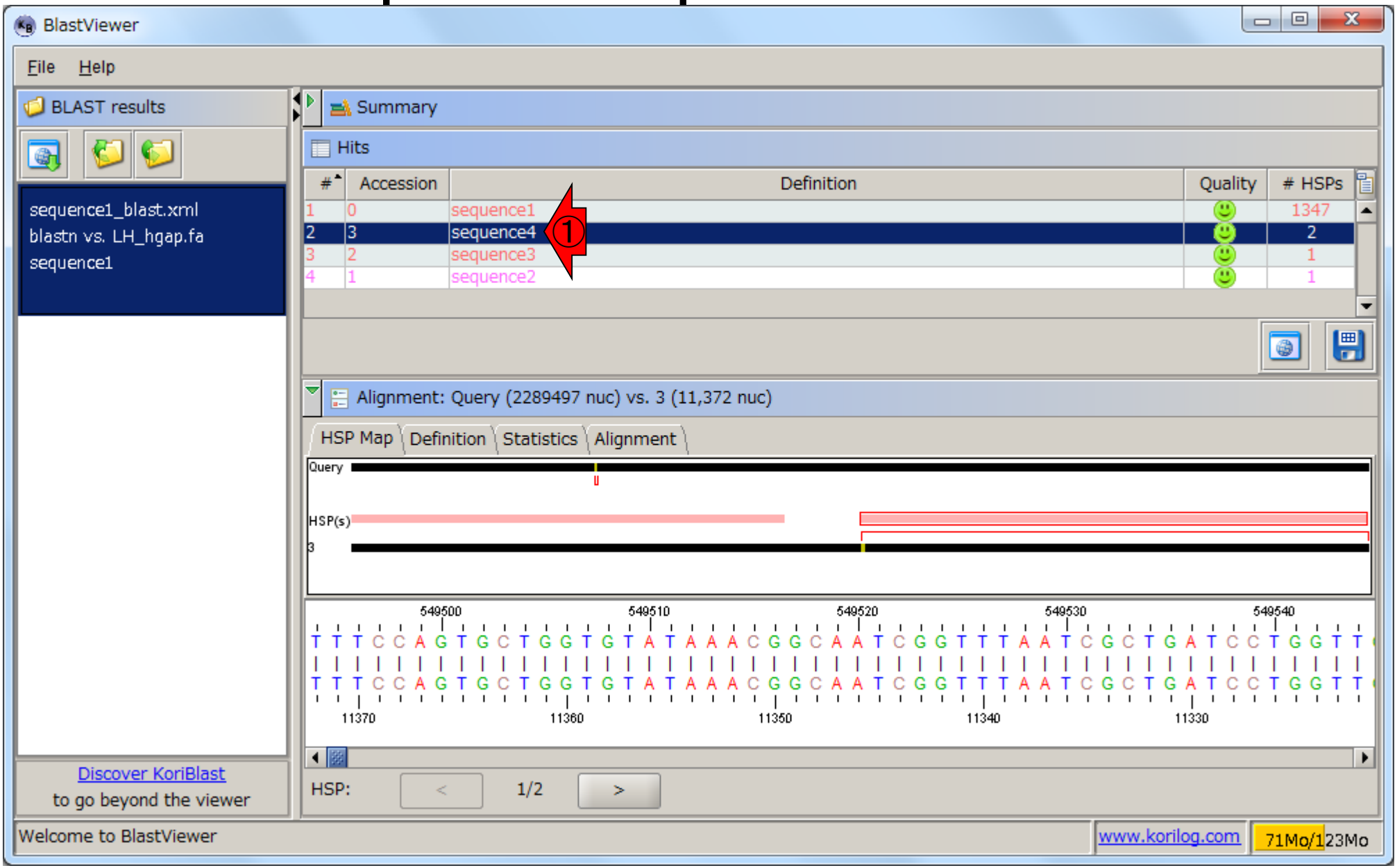
W7-6：grep -n

①grep -nで検索文字列（この場合"Score ="）を含む行番号も表示。②赤枠の上下関係（sequence1の最後の205776とsequence3の206509）から、sequence4とヒットした領域は2か所あり、いずれもスコア（約10320と8907）がそこそこ高い。スコアの合計（2万弱）とsequence4の長さ(11,372 bp)の関係から、sequence4の大部分の領域がsequence1と類似していると判断できる



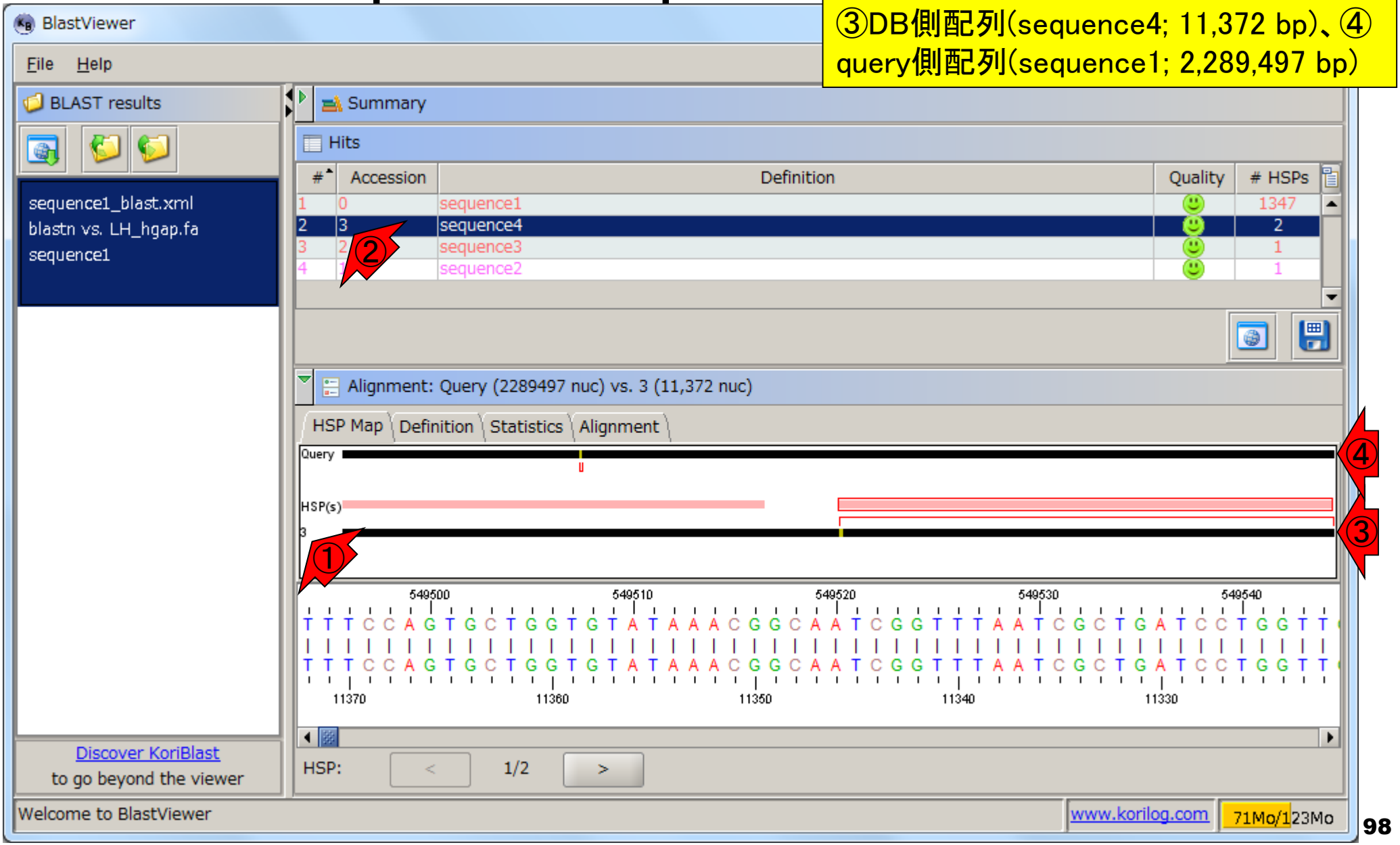
W8-1：seq1 vs. seq4

BlastViewerに戻り、①sequence4とのヒット領域を眺める



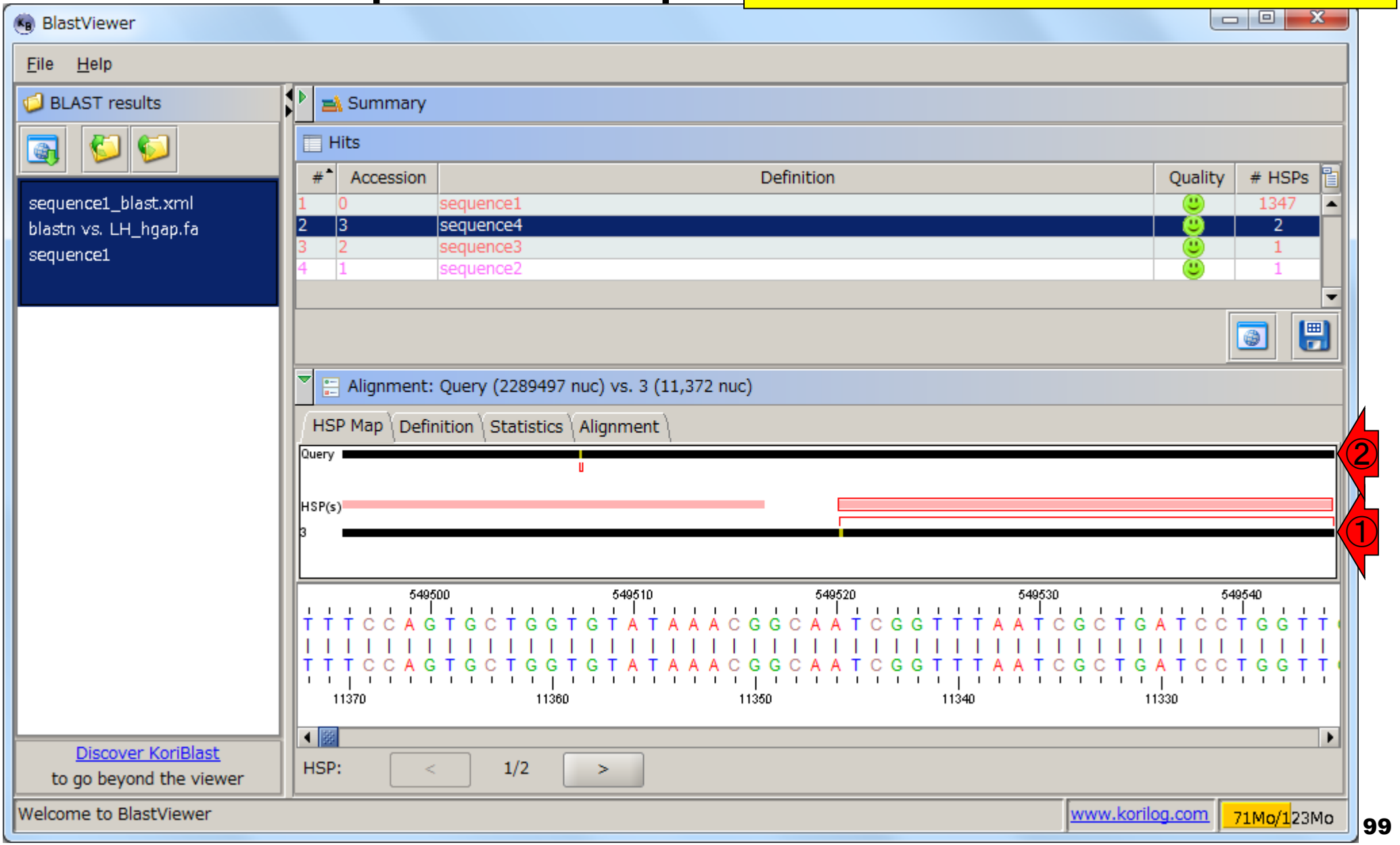
W8-1：seq1 vs. seq4

大まかな見方を説明。①3という数字は、②Accessionのところの数字と同じものであり、3’末端などという意味ではない。③DB側配列(sequence4; 11,372 bp)、④query側配列(sequence1; 2,289,497 bp)



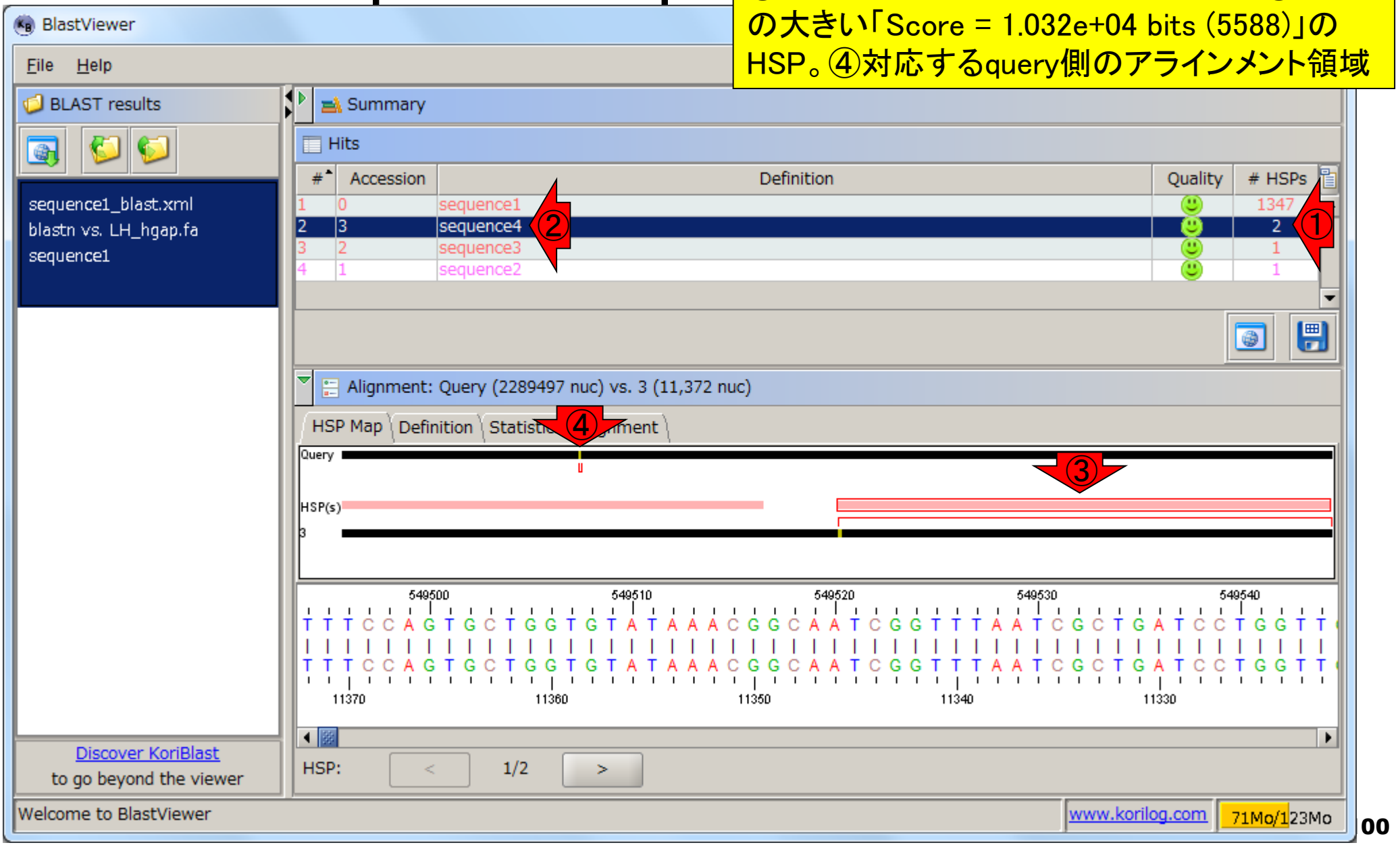
W8-1：seq1 vs. seq4

①DB側配列(sequence4; 11,372 bp)と、②query側配列(sequence1; 2,289,497 bp)は実際の長さは異なるが、長さを揃えて表示している



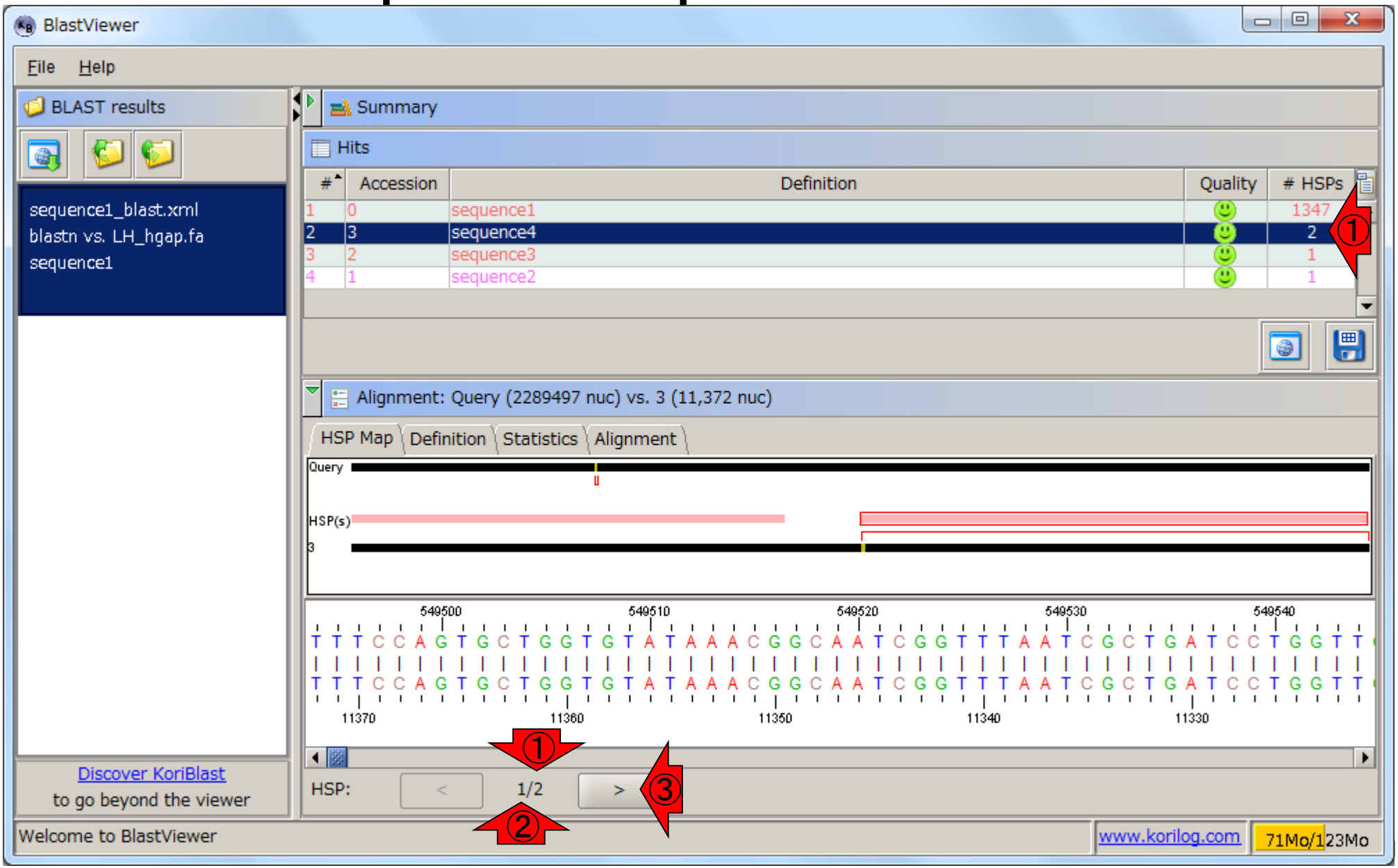
W8-1：seq1 vs. seq4

①(sequence4; 11,372 bp)中には、配列類似領域が2つ(HSPが2個)あったことを思い出そう。
②をクリックして最初に見ているのは、③スコアの大きい「Score = 1.032e+04 bits (5588)」のHSP。④対応するquery側のアラインメント領域



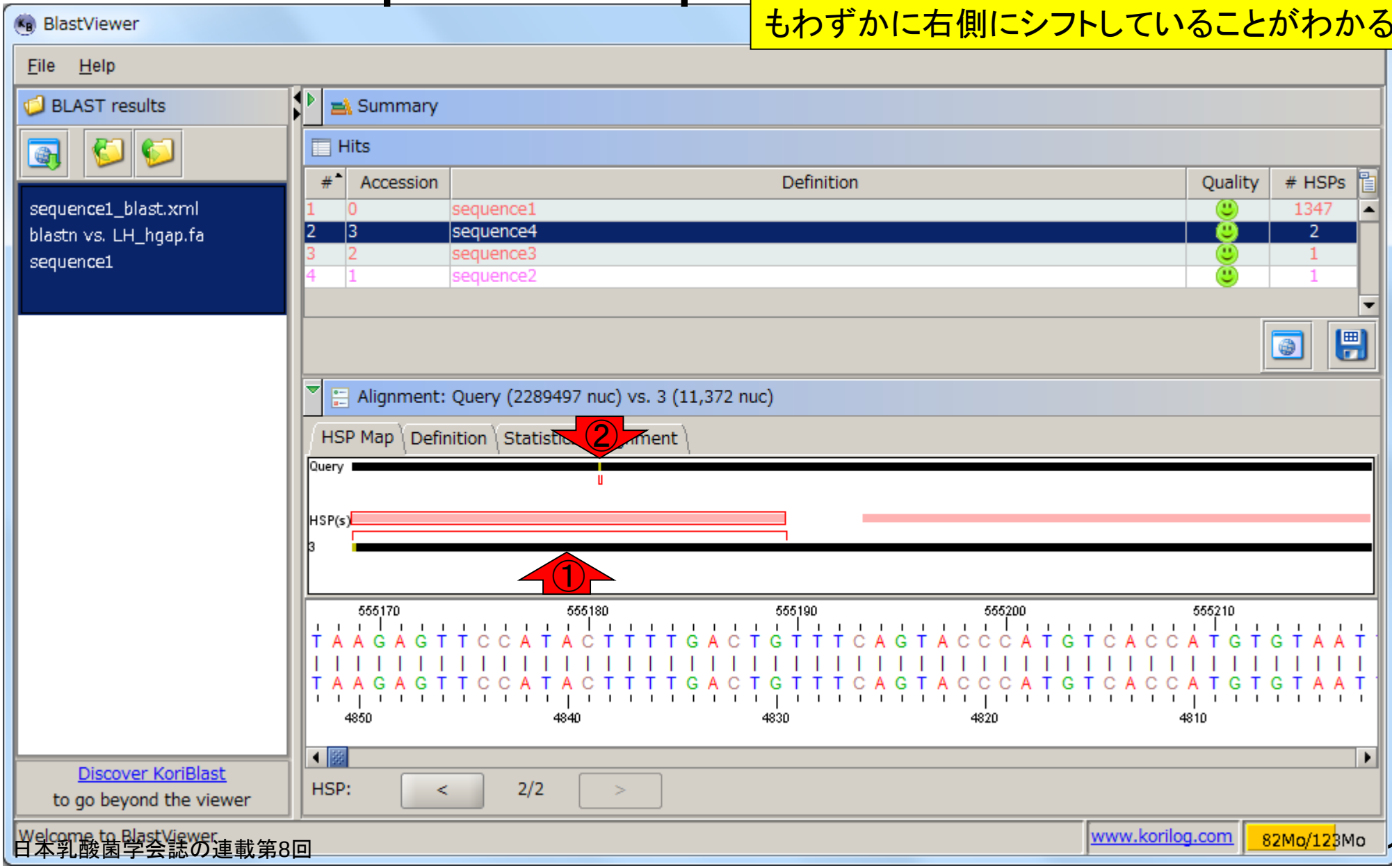
W8-1：seq1 vs. seq4

①2つあるHSPのうち、②最初の1個目という意味。③次のHSPのアラインメントが見られる



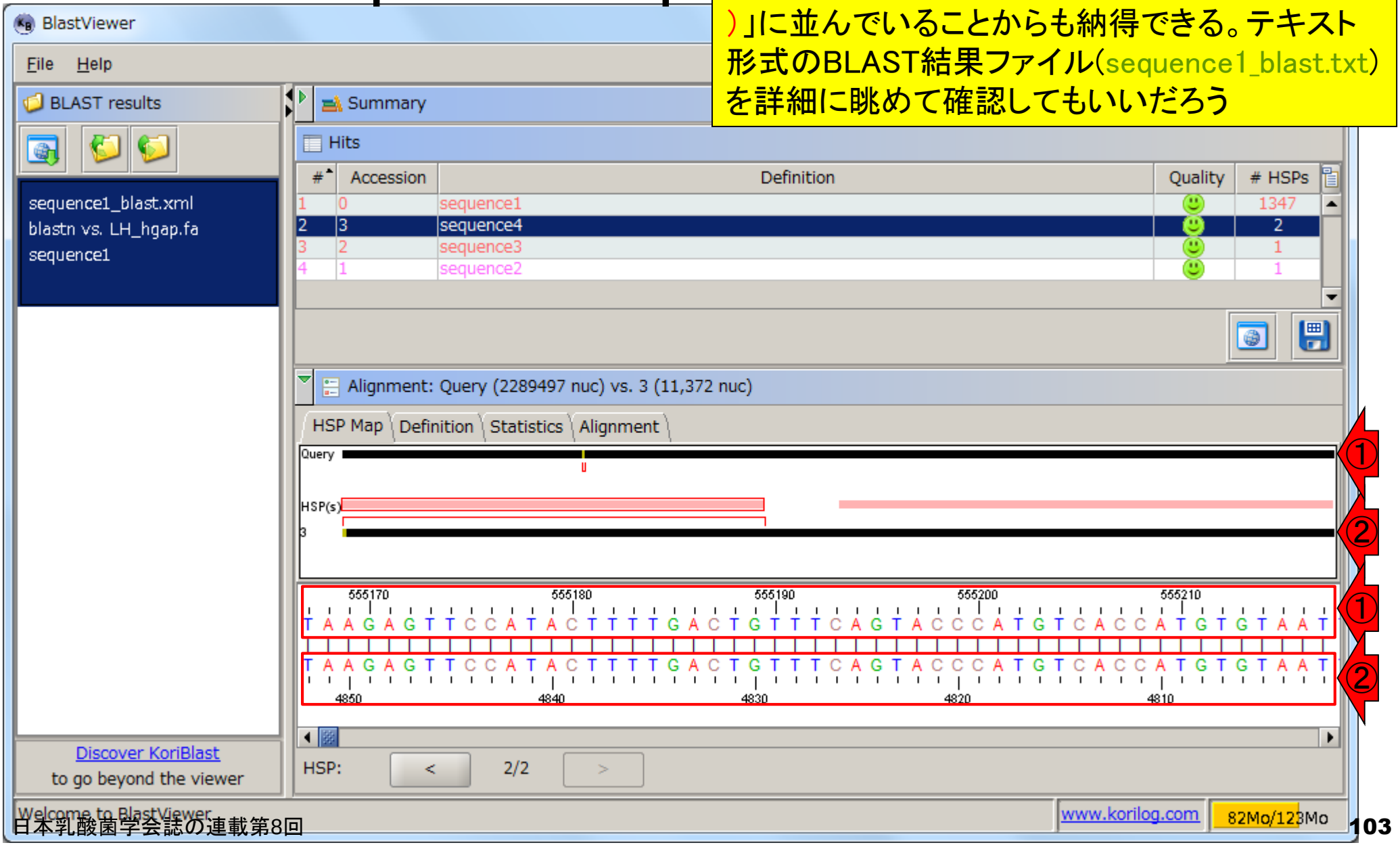
W8-1：seq1 vs. seq4

次のHSP（「Score = 8907 bits (4823)」）のアラインメント。①赤枠の位置が変わっているのがわかる。②対応するquery側のアラインメント領域もわずかに右側にシフトしていることがわかる



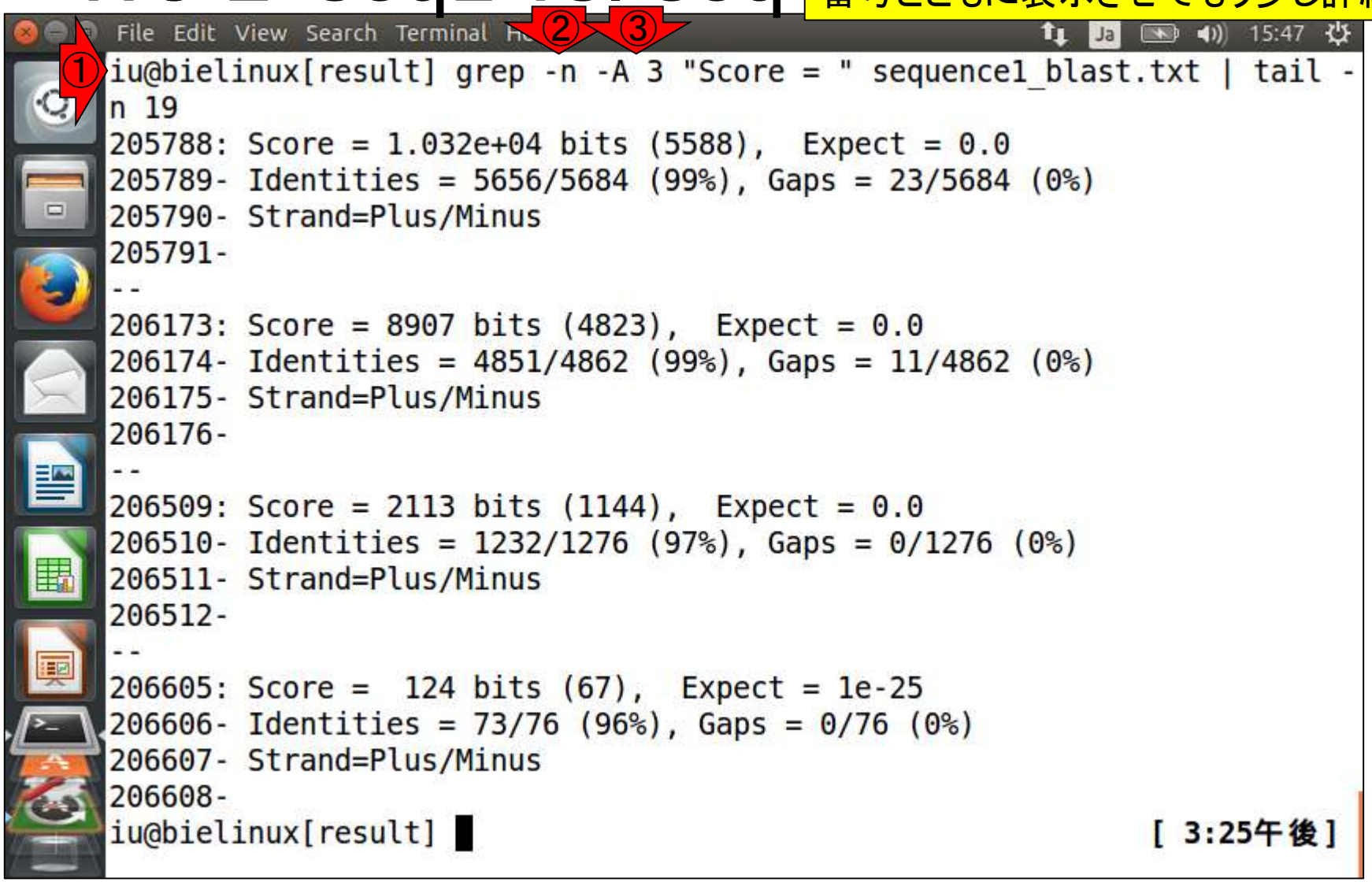
W8-1：seq1 vs. seq4

これは、①query側配列(sequence1)が「左から右の方向(Plus鎖)」に並んでいるのに対して、②DB側配列(sequence4)が「右から左の方向（Minus鎖）」に並んでいることからも納得できる。テキスト形式のBLAST結果ファイル(sequence1_blast.txt)を詳細に眺めて確認してもいいだろう



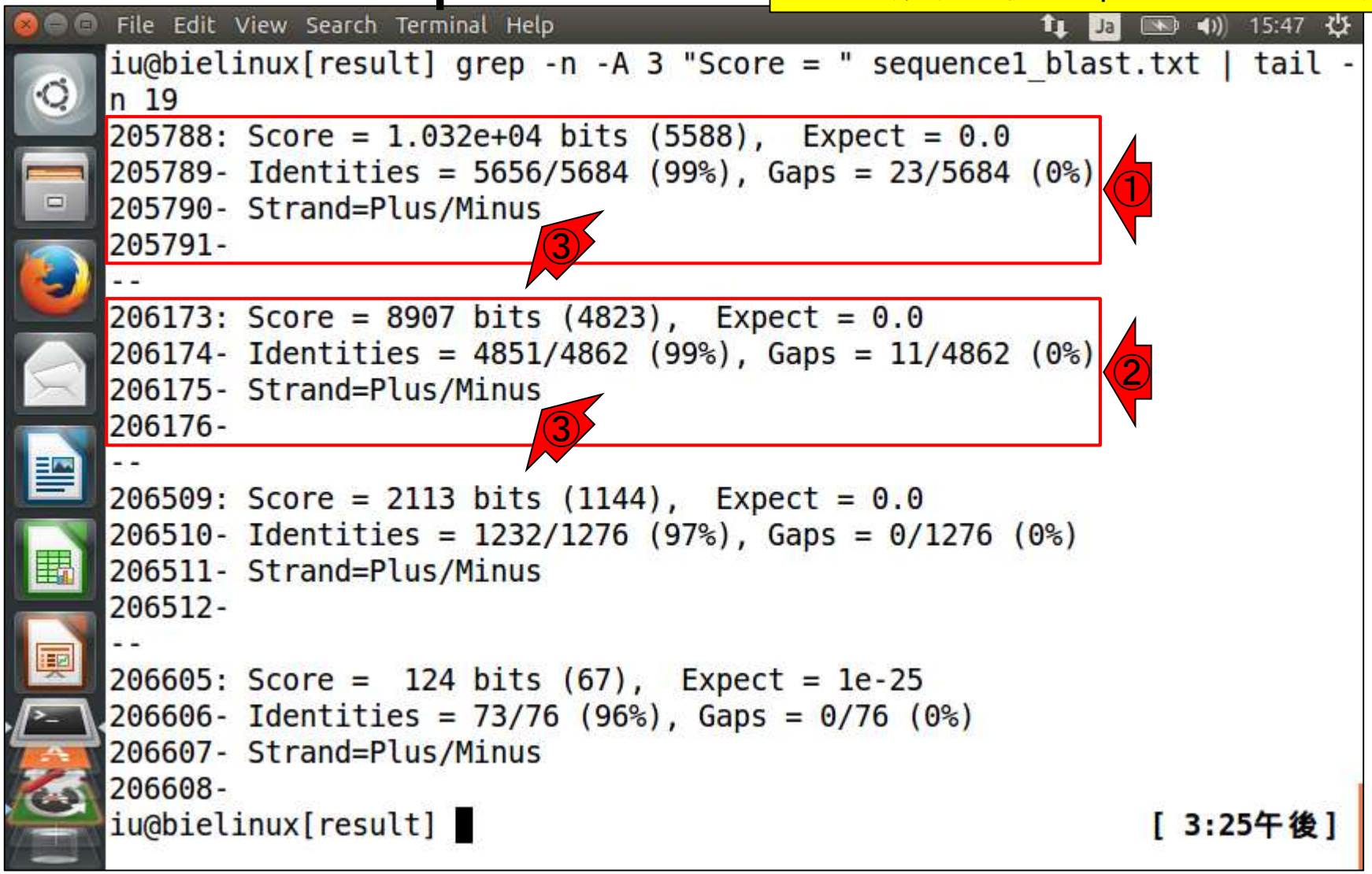
W8-2：seq1 vs. seq4

①grep実行時に②-nと③-A 3を組み合わせて、"Score = "と一致した行を含めた後ろの3行分を行番号とともに表示させてもう少し詳細情報を表示



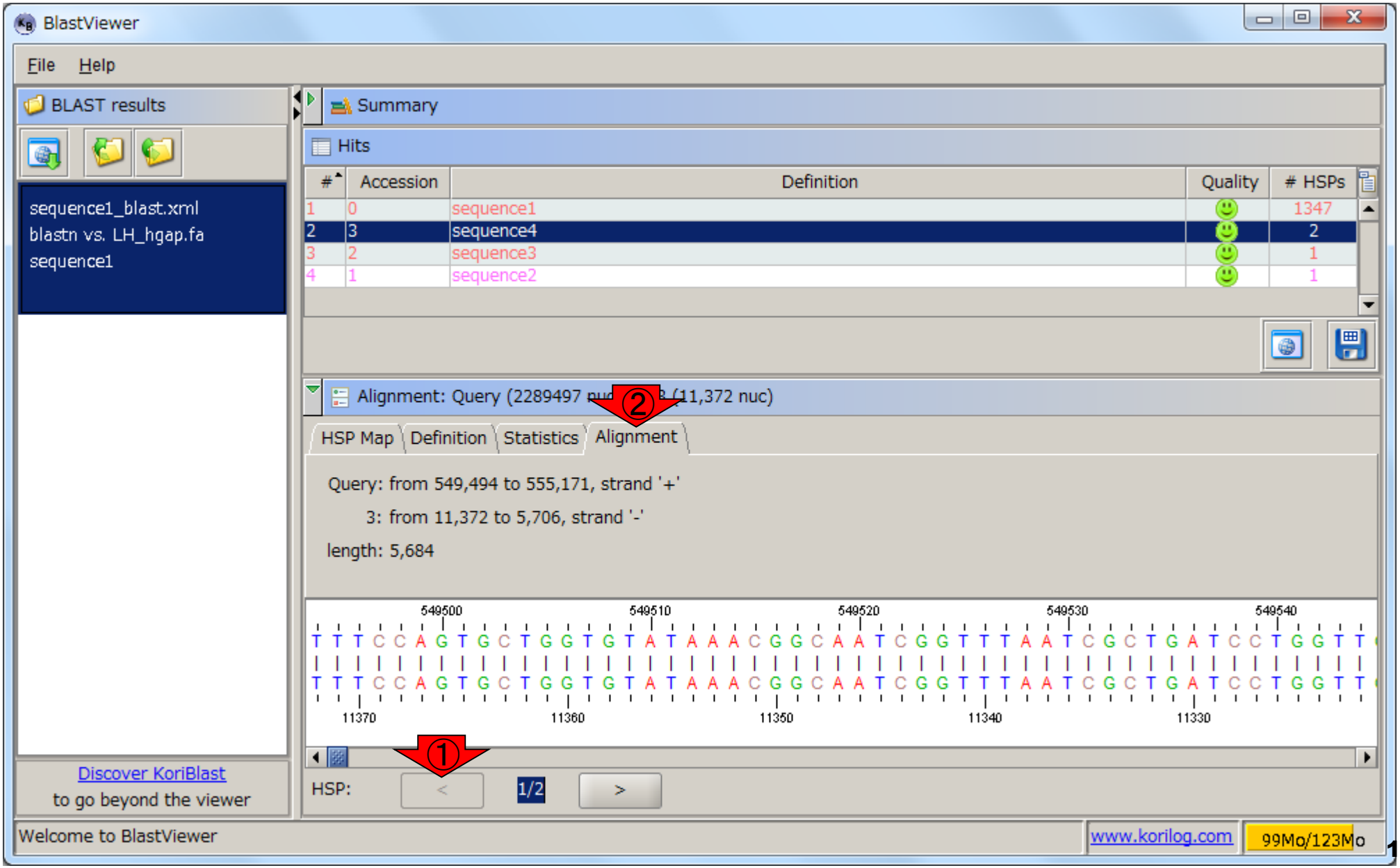
W8-2：seq1 vs. seq

sequence4上の配列類似領域の①1つめと②2つめ。
③確かにストランド情報として、query側(sequence1)がPlus鎖、DB側(sequence4)がMinus鎖となっている



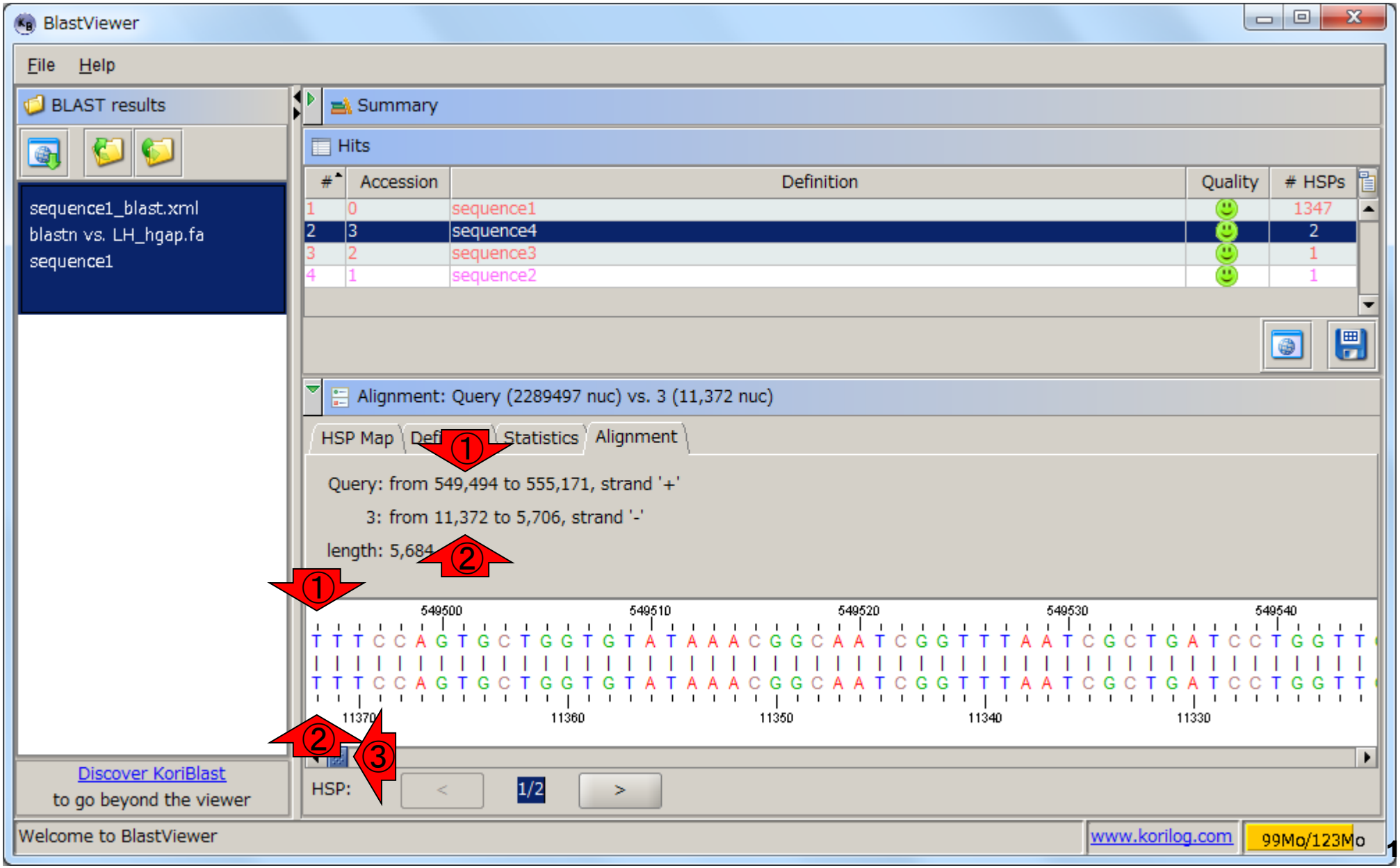
W8-3：HSP1

①1つめのHSP（以下、HSP1)に切り替えて、②Alignmentタブをクリック。ここでアラインメント結果の詳細情報がわかる



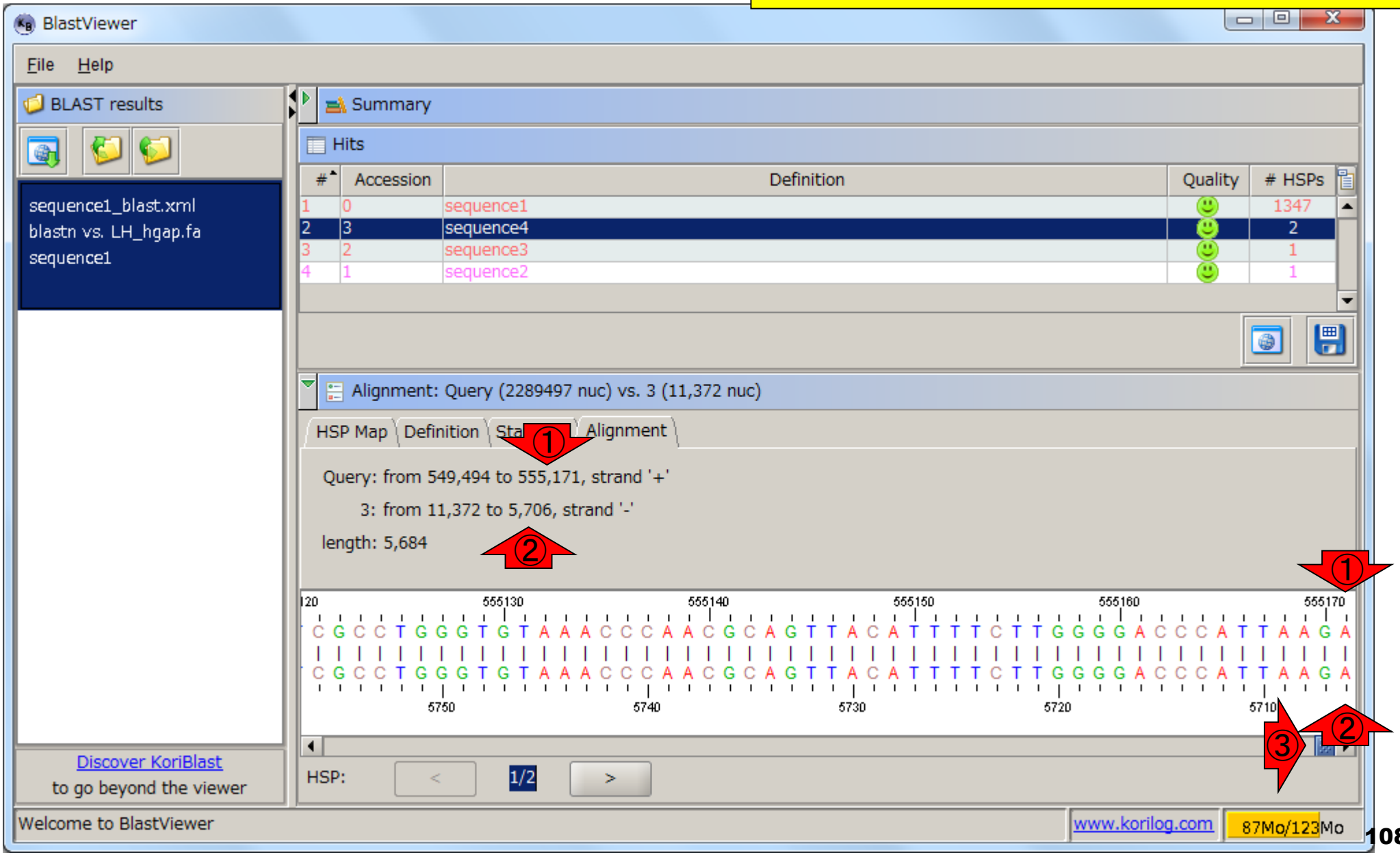
W8-3：HSP1

①query側(sequence1)の549,494番目と、②DB側(sequence4)の11,372番目から…③を一番右まで移動



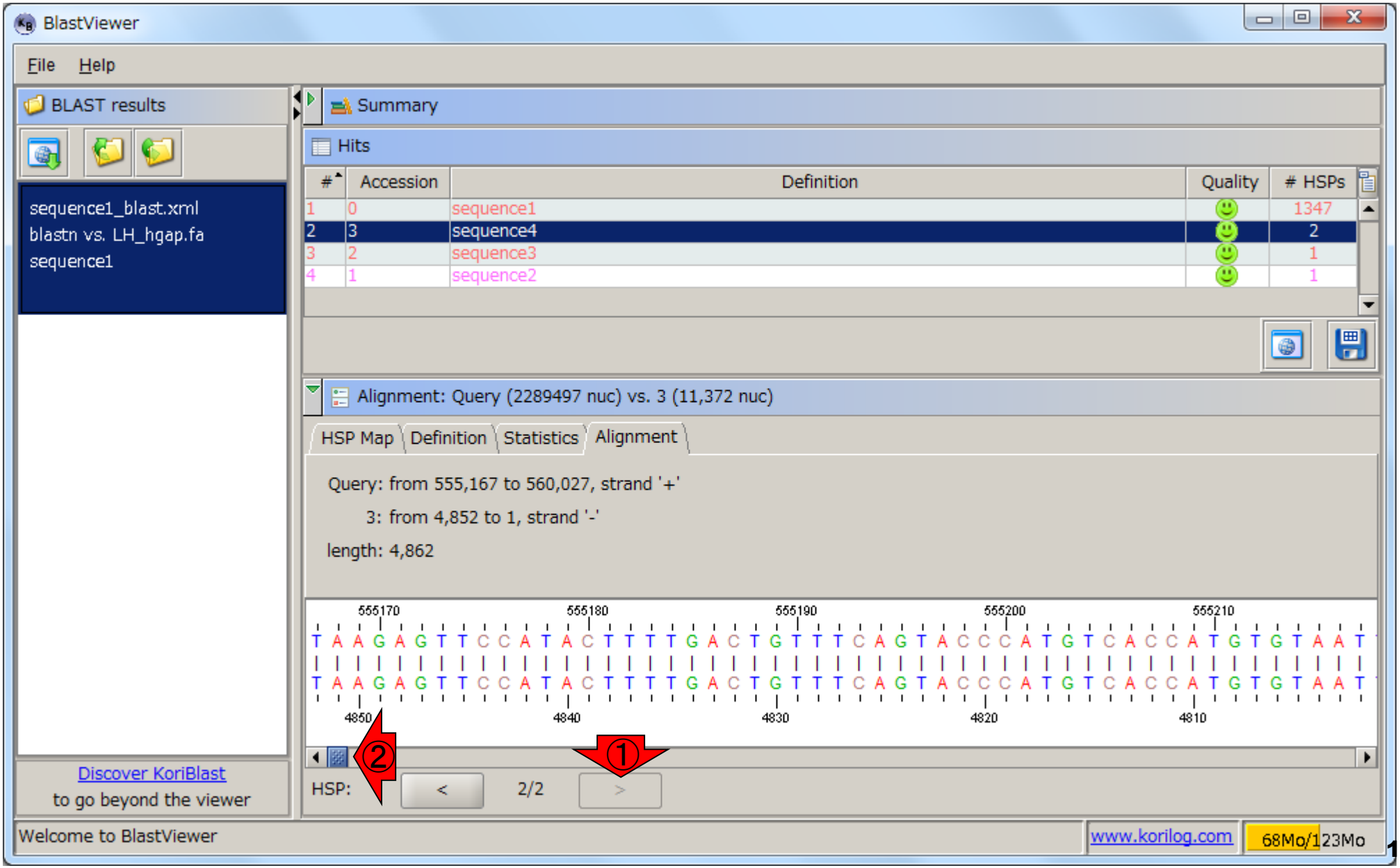
W8-3：HSP1

①query側(sequence1)の555,171番目と、②DB側(sequence4)の5,706番目の領域がHSP1。③を一番右まで移動させた結果で見ています



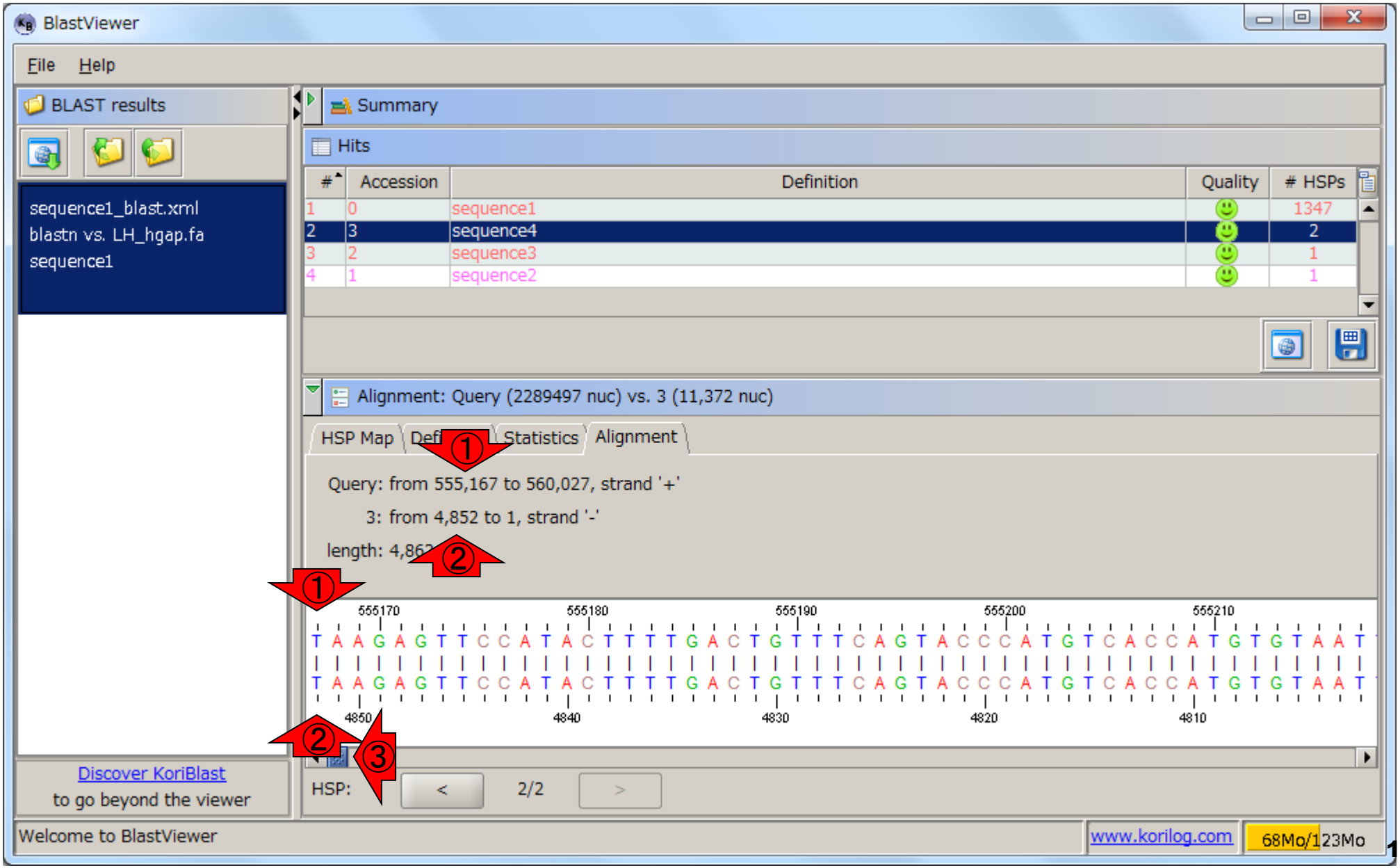
①2つめのHSP（以下、HSP2)に切り替えて、②一番左に移動

W8-4：HSP2



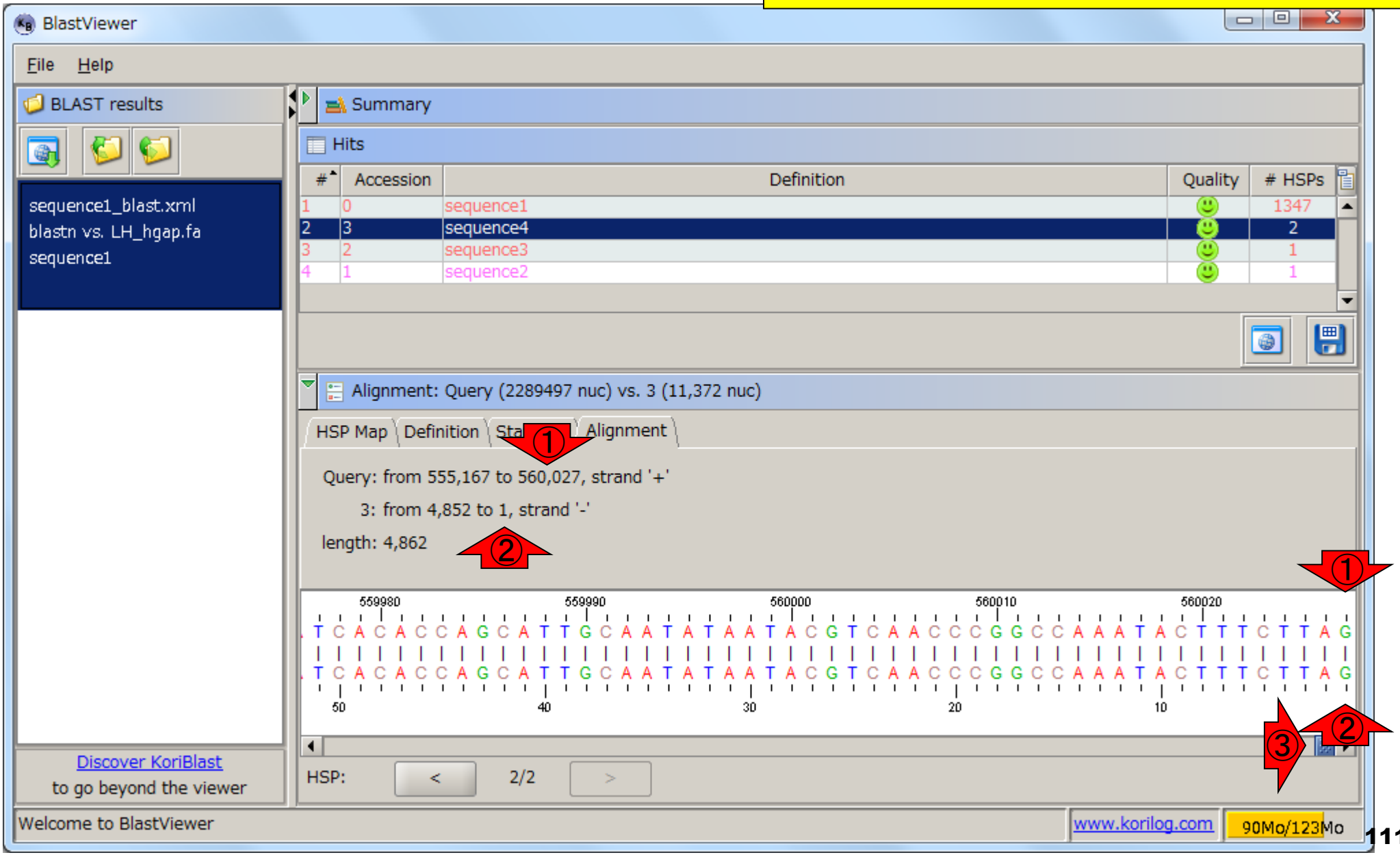
W8-4：HSP2

①query側(sequence1)の555,167番目と、②DB側(sequence4)の4,852番目から…③を一番右まで移動



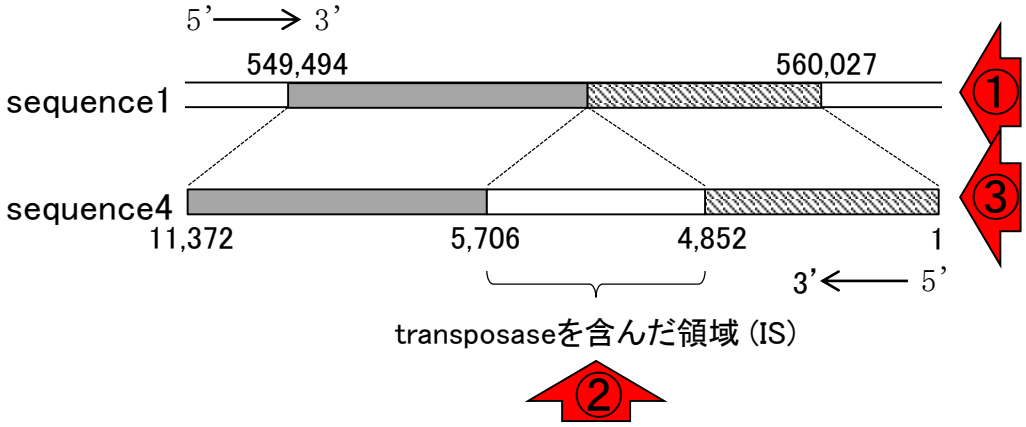
W8-4：HSP2

①query側(sequence1)の560,027番目と、②DB側(sequence4)の1番目の領域がHSP2。③を一番右まで移動させた結果で見ています



W8-5：模式図

sequence1とsequence4のアラインメント模式図。①sequence1の一続きの領域[549494, 560027 bp]と、②領域[4853, 5705 bp]を除く、③sequence4の全長がほぼ一致。そして、④アラインメントされなかった領域[4853, 5705 bp]には、transposaseがコードされていた(W4-9)。第8回の図1と同じ

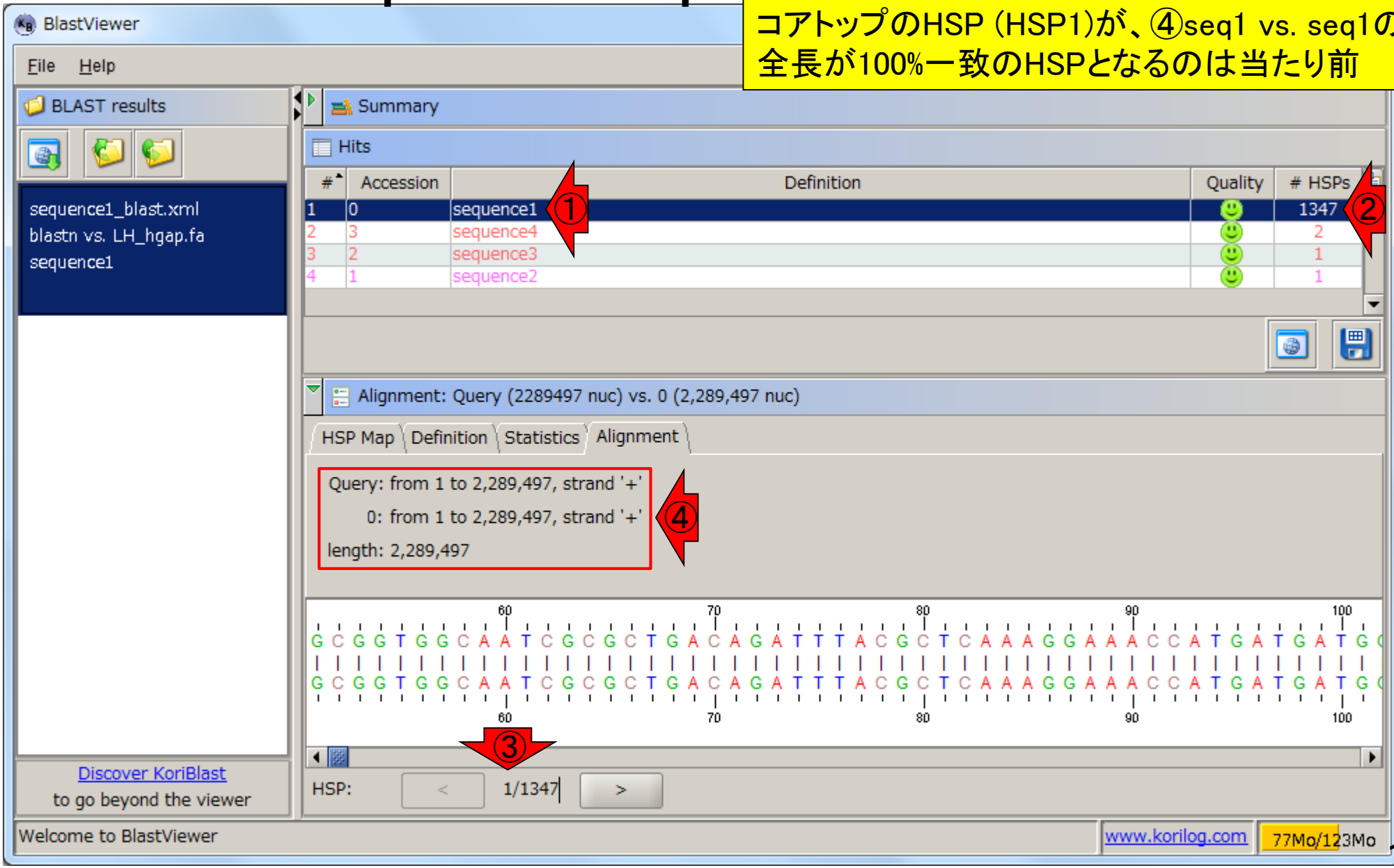


sequence4	4133..4753	CDS	mannose-specific PTS system IID component
sequence4	4781..4873	CDS	mannose-specific PTS system IIA component
sequence4	4912..5220	CDS	transposase
sequence4	5361..5699	CDS	transposase
sequence4	5711..6058	CDS	mannose-specific PTS system IIA component
sequence4	6114..6596	CDS	mannose/fructose/sorbose-specific PTS system IID component

④

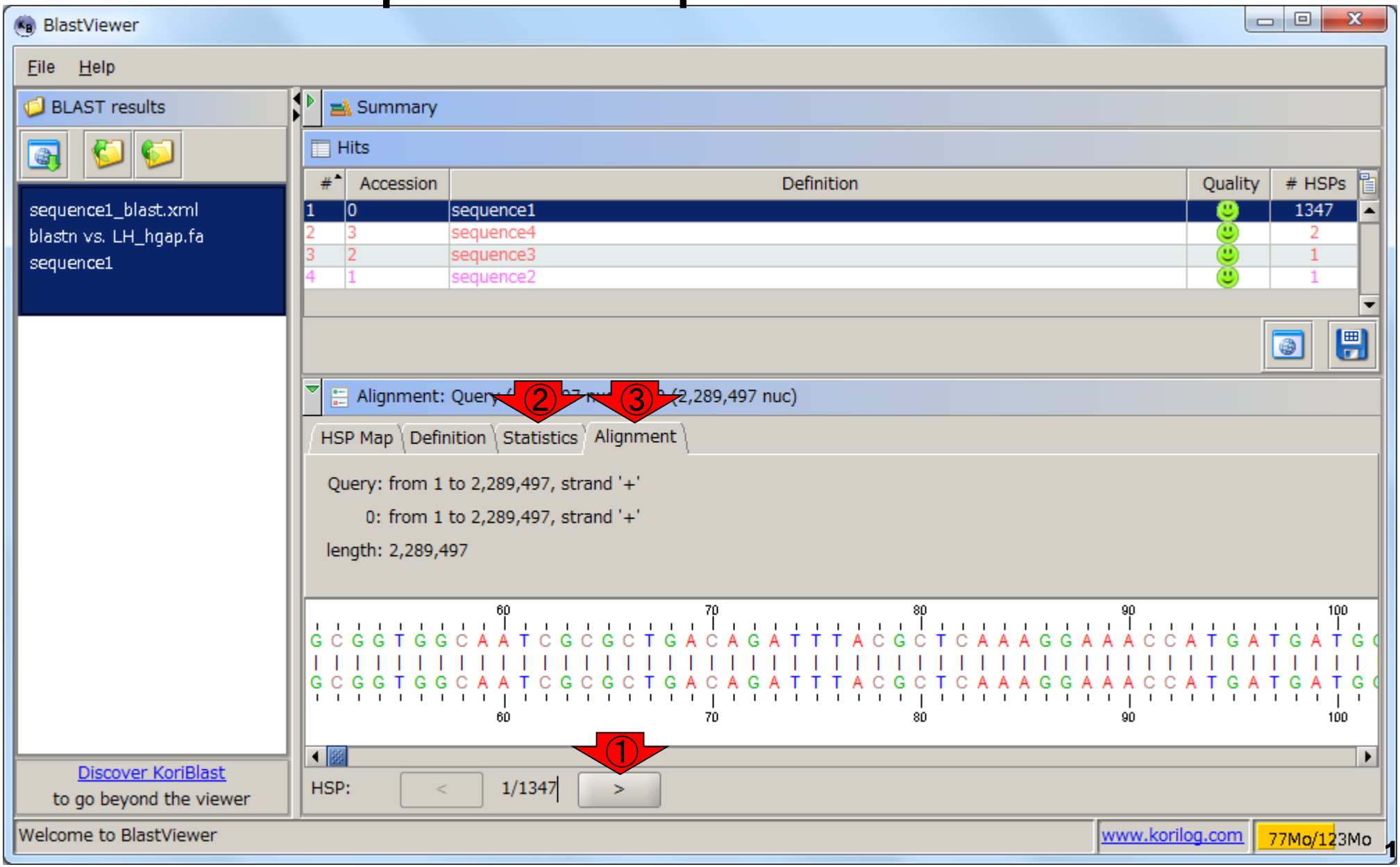
W9-1：seq1 vs. seq1

query側(sequence1)は固定で、①DB側がsequence1のBLAST結果を表示。②配列類似領域(HSP)は1,347個あったことを思い出そう。③スコアトップのHSP (HSP1)が、④seq1 vs. seq1の全長が100%一致のHSPとなるのは当たり前



W9-1：seq1 vs. seq1

①や②や③を押して、スコアの上位33位(HSP1−33)までの全体像をまとめたのが…



W9-2：HSP1-33

これです。①括弧内のHSPは、カッコ外のHSPとqueryとDB側の一致領域が入れ替わっているだけで、実質的に同じものです

HSP番号	スコア	一致領域(query側)			一致領域(DB側)			Length	query側の領域	DB側の領域
		start	end	strand	start	end	strand			
HSP1	4,227,900	1	2,289,497	plus	1	2,289,497	plus	2,289,497		
HSP2 (HSP3)	11,624	2,058,465	2,065,193	plus	2,057,850	2,064,578	plus	6,732		
HSP4 (HSP5)	10,754	37,329	43,187	plus	1	5,860	plus	5,865	①'	①
HSP6 (HSP7)	10,516	2,059,080	2,065,193	plus	2,057,850	2,063,963	plus	6,117		
HSP8 (HSP9)	10,429	2,283,820	2,289,497	plus	5,839	11,509	plus	5,679	②'	②
HSP10 (HSP11)	9,679	2,273,804	2,279,100	plus	717,190	711,892	minus	5,301		
HSP12 (HSP13)	9,524	2,059,672	2,065,193	plus	2,057,830	2,063,348	plus	5,525		
HSP14 (HSP15)	9,362	1,088,170	1,093,274	plus	999,673	1,004,778	plus	5,106		
HSP16 (HSP17)	8,385	2,060,310	2,065,194	plus	2,057,850	2,062,735	plus	4,887		
HSP18 (HSP19)	7,301	2,060,901	2,065,194	plus	2,057,830	2,062,120	plus	4,296		
HSP20 (HSP21)	6,186	2,061,516	2,065,193	plus	2,057,830	2,061,504	plus	3,680		
HSP22 (HSP23)	6,047	2,275,808	2,279,168	plus	1,002,969	999,608	minus	3,365		
HSP24 (HSP**26**)	6,043	999,670	1,002,969	plus	711,886	715,184	plus	3,300		
HSP25 (HSP**27**)	6,043	1,088,173	1,091,466	plus	711,892	715,184	plus	3,294		
HSP28 (HSP29)	6,038	2,275,808	2,279,100	plus	1,091,466	1,088,173	minus	3,295		
HSP30 (HSP31)	5,050	2,062,154	2,065,193	plus	2,057,850	2,060,889	plus	3,042		
HSP32 (HSP33)	4,021	2,062,769	2,065,197	plus	2,057,850	2,060,279	plus	2,431		

①の数値でソートした結果

W9-2：HSP1-33

HSP番号	スコア	一致領域(query側)			一致領域(DB側)			Length	query側の領域	DB側の領域
		start	end	strand	start	end	strand			
HSP1	4,227,900	1	2,289,497	plus	1	2,289,497	plus	2,289,497		
HSP4 (HSP5)	10,754	37,329	43,187	plus	1	5,860	plus	5,865	①'	①
HSP24 (HSP**26**)	6,043	999,670	1,002,969	plus	711,886	715,184	plus	3,300		
HSP14 (HSP15)	9,362	1,088,170	1,093,274	plus	999,673	1,004,778	plus	5,106		
HSP25 (HSP**27**)	6,043	1,088,173	1,091,466	plus	711,892	715,184	plus	3,294		
HSP2 (HSP3)	11,624	2,058,465	2,065,193	plus	2,057,850	2,064,578	plus	6,732		
HSP6 (HSP7)	10,516	2,059,080	2,065,193	plus	2,057,850	2,063,963	plus	6,117		
HSP12 (HSP13)	9,524	2,059,672	2,065,193	plus	2,057,830	2,063,348	plus	5,525		
HSP16 (HSP17)	8,385	2,060,310	2,065,194	plus	2,057,850	2,062,735	plus	4,887		
HSP18 (HSP19)	7,301	2,060,901	2,065,194	plus	2,057,830	2,062,120	plus	4,296		
HSP20 (HSP21)	6,186	2,061,516	2,065,193	plus	2,057,830	2,061,504	plus	3,680		
HSP30 (HSP31)	5,050	2,062,154	2,065,193	plus	2,057,850	2,060,889	plus	3,042		
HSP32 (HSP33)	4,021	2,062,769	2,065,197	plus	2,057,850	2,060,279	plus	2,431		
HSP10 (HSP11)	9,679	2,273,804	2,279,100	plus	717,190	711,892	minus	5,301		
HSP22 (HSP23)	6,047	2,275,808	2,279,168	plus	1,002,969	999,608	minus	3,365		
HSP28 (HSP29)	6,038	2,275,808	2,279,100	plus	1,091,466	1,088,173	minus	3,295		
HSP8 (HSP9)	10,429	2,283,820	2,289,497	plus	5,839	11,509	plus	5,679	②'	②

①や②の一致領域が、後の議論対象です

W9-2：HSP1-33

HSP番号	スコア	一致領域(query側)			一致領域(DB側)			Length	query側の領域	DB側の領域
		start	end	strand	start	end	strand			
HSP1	4,227,900	1	2,289,497	plus	1	2,289,497	plus	2,289,497		
HSP4 (HSP5)	10,7	37,329	43,187	plus	1	5,860	plus	5,865	①'	①
HSP24 (HSP26)	6,04	999,670	1,002,969	plus	711,886	715,184	plus	3,300		
HSP14 (HSP15)	9,362	1,088,170	1,093,274	plus	999,673	1,004,778	plus	5,106		
HSP25 (HSP27)	6,043	1,088,173	1,091,466	plus	711,892	715,184	plus	3,294		
HSP2 (HSP3)	11,624	2,058,465	2,065,193	plus	2,057,850	2,064,578	plus	6,732		
HSP6 (HSP7)	10,516	2,059,080	2,065,193	plus	2,057,850	2,063,963	plus	6,117		
HSP12 (HSP13)	9,524	2,059,672	2,065,193	plus	2,057,830	2,063,348	plus	5,525		
HSP16 (HSP17)	8,385	2,060,310	2,065,194	plus	2,057,850	2,062,735	plus	4,887		
HSP18 (HSP19)	7,301	2,060,901	2,065,194	plus	2,057,830	2,062,120	plus	4,296		
HSP20 (HSP21)	6,186	2,061,516	2,065,193	plus	2,057,830	2,061,504	plus	3,680		
HSP30 (HSP31)	5,050	2,062,154	2,065,193	plus	2,057,850	2,060,889	plus	3,042		
HSP32 (HSP33)	4,021	2,062,769	2,065,197	plus	2,057,850	2,060,279	plus	2,431		
HSP10 (HSP11)	9,679	2,273,804	2,279,100	plus	717,190	711,892	minus	5,301		
HSP22 (HSP23)	6,047	2,275,808	2,279,168	plus	1,002,969	999,608	minus	3,365		
HSP28 (HSP29)	6,03	2,275,808	2,279,100	plus	1,091,466	1,088,173	minus	3,295		
HSP8 (HSP9)	10,4	2,283,820	2,289,497	plus	5,839	11,509	plus	5,679	②'	②

W9-2：HSP1-33

右側の①’が領域[37329, 43187 bp]、①が[1, 5860 bp] 、②’が [2283820, 2289497 bp]、②が[5839, 11509 bp]に相当。ここの「①’, ①, ②’, ②」が次のスライドで使われる

HSP番号	スコア	一致領域(query側)			一致領域(DB側)			Length	query側の領域	DB側の領域
		start	end	strand	start	end	strand			
HSP1	4,227,900	1	2,289,497	plus	1	2,289,497	plus	2,289,497		
HSP4 (HSP5)	10,754	37,329	43,187	plus	1	5,860	plus	5,865	①’	①
HSP24 (HSP**26**)	6,043	999,670	1,002,969	plus	711,886	715,184	plus	3,300		
HSP14 (HSP15)	9,362	1,088,170	1,093,274	plus	999,673	1,004,778	plus	5,106		
HSP25 (HSP**27**)	6,043	1,088,173	1,091,466	plus	711,892	715,184	plus	3,294		
HSP2 (HSP3)	11,624	2,058,465	2,065,193	plus	2,057,850	2,064,578	plus	6,732		
HSP6 (HSP7)	10,516	2,059,080	2,065,193	plus	2,057,850	2,063,963	plus	6,117		
HSP12 (HSP13)	9,524	2,059,672	2,065,193	plus	2,057,830	2,063,348	plus	5,525		
HSP16 (HSP17)	8,385	2,060,310	2,065,194	plus	2,057,850	2,062,735	plus	4,887		
HSP18 (HSP19)	7,301	2,060,901	2,065,194	plus	2,057,830	2,062,120	plus	4,296		
HSP20 (HSP21)	6,186	2,061,516	2,065,193	plus	2,057,830	2,061,504	plus	3,680		
HSP30 (HSP31)	5,050	2,062,154	2,065,193	plus	2,057,850	2,060,889	plus	3,042		
HSP32 (HSP33)	4,021	2,062,769	2,065,197	plus	2,057,850	2,060,279	plus	2,431		
HSP10 (HSP11)	9,679	2,273,804	2,279,100	plus	717,190	711,892	minus	5,301		
HSP22 (HSP23)	6,047	2,275,808	2,279,168	plus	1,002,969	999,608	minus	3,365		
HSP28 (HSP29)	6,038	2,275,808	2,279,100	plus	1,091,466	1,088,173	minus	3,295		
HSP8 (HSP9)	10,429	2,283,820	2,289,497	plus	5,839	11,509	plus	5,679	②’	②

③のあたりはribosomal RNAがほとんど

W9-3：アノテーション

HSP番号	スコア	一致領域(query側)			一致領域(DB側)			Length	query側の領域	DB側の領域
		start	end	strand	start	end	strand			
HSP1	4,227,900	1	2,289,497	plus	1	2,289,497	plus	2,289,497		
HSP4 (HSP5)	10,754	37,329	43,187	plus	1	5,860	plus	5,865	①'	①
HSP24 (HSP26)	6,04[illegible]	999,670	1,002,969	plus	711,886	715,184	plus	3,300	ribosomal RNA	
HSP14 (HSP15)	9,3[illegible]	1,088,170	1,093,274	plus	999,673	1,004,778	plus	5,106	ribosomal RNA	
HSP25 (HSP27)	6,04[illegible]	1,088,173	1,091,466	plus	711,892	715,184	plus	3,294	ribosomal RNA	
HSP2 (HSP3)	11,624	2,058,465	2,065,193	plus	2,057,850	2,064,578	plus	6,732		
HSP6 (HSP7)	10,516	2,059,080	2,065,193	plus	2,057,850	2,063,963	plus	6,117		
HSP12 (HSP13)	9,524	2,059,672	2,065,193	plus	2,057,830	2,063,348	plus	5,525		
HSP16 (HSP17)	8,385	2,060,310	2,065,194	plus	2,057,850	2,062,735	plus	4,887		
HSP18 (HSP19)	7,301	2,060,901	2,065,194	plus	2,057,830	2,062,120	plus	4,296		
HSP20 (HSP21)	6,186	2,061,516	2,065,193	plus	2,057,830	2,061,504	plus	3,680		
HSP30 (HSP31)	5,050	2,062,154	2,065,193	plus	2,057,850	2,060,889	plus	3,042		
HSP32 (HSP33)	4,021	2,062,769	2,065,197	plus	2,057,850	2,060,279	plus	2,431		
HSP10 (HSP11)	9,67[illegible]	2,273,804	2,279,100	plus	717,190	711,892	minus	5,301	ribosomal RNA	
HSP22 (HSP23)	6,0[illegible]	2,275,808	2,279,168	plus	1,002,969	999,608	minus	3,365	ribosomal RNA	
HSP28 (HSP29)	6,03[illegible]	2,275,808	2,279,100	plus	1,091,466	1,088,173	minus	3,295	ribosomal RNA	
HSP8 (HSP9)	10,429	2,283,820	2,289,497	plus	5,839	11,509	plus	5,679	②'	②

W9-3：アノテーション

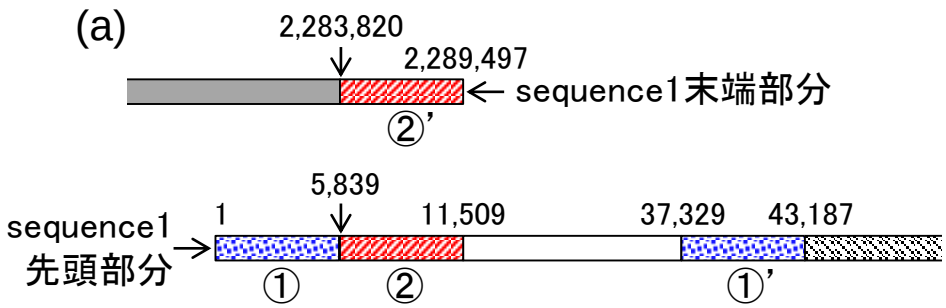
④のあたりは、「adhesion exoprotein、mucus-binding protein、hypothetical protein」となっており、はっきり言ってよくわからない

HSP番号	スコア	一致領域(query側) start	end	strand	一致領域(DB側) start	end	strand	Length	query側の領域	DB側の領域
HSP1	4,227,900	1	2,289,497	plus	1	2,289,497	plus	2,289,497		
HSP4 (HSP5)	10,754	37,329	43,187	plus	1	5,860	plus	5,865	①'	①
HSP24 (HSP26)	6,043	999,670	1,002,969	plus	711,886	715,184	plus	3,300	ribosomal RNA	
HSP14 (HSP15)	9,362	1,088,170	1,093,274	plus	999,673	1,004,778	plus	5,106	ribosomal RNA	
HSP25 (HSP27)	6,043	1,088,173	1,091,466	plus	711,892	715,184	plus	3,294	ribosomal RNA	
HSP2 (HSP3)	11,624	2,058,465	2,065,193	plus	2,057,850	2,064,578	plus	6,732		
HSP6 (HSP7)	10,516	2,059,080	2,065,193	plus	2,057,850	2,063,963	plus	6,117		
HSP12 (HSP13)	9,524	2,059,672	2,065,193	plus	2,057,830	2,063,348	plus	5,525		
HSP16 (HSP17)	8,38	2,060,310	2,065,194	plus	2,057,850	2,062,735	plus	4,887		
HSP18 (HSP19)	7,30	2,060,901	2,065,194	plus	2,057,830	2,062,120	plus	4,296		
HSP20 (HSP21)	6,186	2,061,516	2,065,193	plus	2,057,830	2,061,504	plus	3,680		
HSP30 (HSP31)	5,050	2,062,154	2,065,193	plus	2,057,850	2,060,889	plus	3,042		
HSP32 (HSP33)	4,021	2,062,769	2,065,197	plus	2,057,850	2,060,279	plus	2,431		
HSP10 (HSP11)	9,679	2,273,804	2,279,100	plus	717,190	711,892	minus	5,301	ribosomal RNA	
HSP22 (HSP23)	6,047	2,275,808	2,279,168	plus	1,002,969	999,608	minus	3,365	ribosomal RNA	
HSP28 (HSP29)	6,038	2,275,808								NA
HSP8 (HSP9)	10,429	2,283,820								②

④

sequence1	2055735..2060864	CDS	adhesion exoprotein
sequence1	2060942..2063323	CDS	mucus-binding protein
sequence1	2063401..2066928	CDS	hypothetical protein

W10-1：sequence1

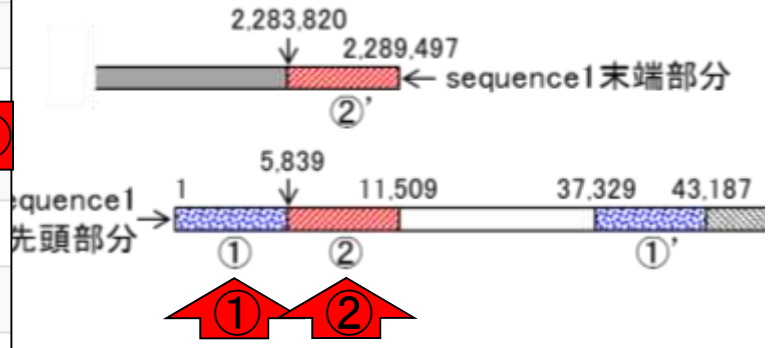


sequence1の末端付近の重複は、HSP8の一致領域（②と②'）に相当。これにHSP4の一致領域（①と①'）を含めた模式図。第8回の図2aと同じ。後の検証により、①の領域は染色体上に存在しないことが確認された

W10-2：アノテーション

HSP4の一致領域①[1, 5860 bp]のDFASTアノテーション。HSP8の一致領域②[5839, 11509 bp]のアノテーションの一部

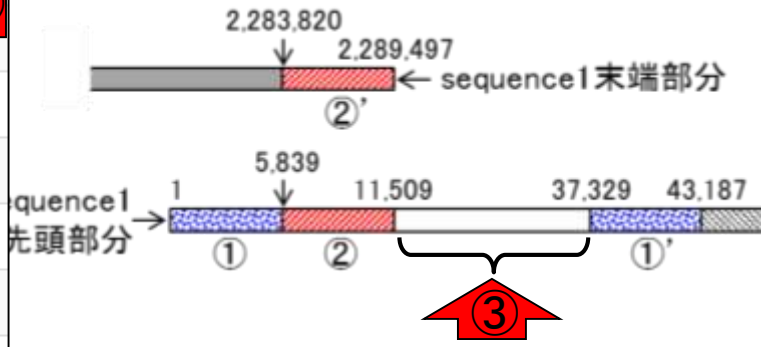
LOCUS_00001	sequence1	151..384	CDS	hypothetical protein
LOCUS_00002	sequence1	350..886	CDS	hypothetical protein
LOCUS_00003	sequence1	883..1311	CDS	hypothetical protein
LOCUS_00004	sequence1	1637..1849	CDS	hypothetical protein
LOCUS_00005	sequence1	1968..2165	CDS	hypothetical protein
LOCUS_00006	sequence1	2355..2732	CDS	prophage protein
LOCUS_00007	sequence1	2725..2940	CDS	hypothetical protein
LOCUS_00008	sequence1	2930..3031	CDS	hypothetical protein
LOCUS_00009	sequence1	3067..3534	CDS	holin
LOCUS_00010	sequence1	3547..4578	CDS	1,4-beta-N-acetylmuramidase
LOCUS_00011	sequence1	4765..4935	CDS	hypothetical protein
LOCUS_00012	sequence1	4936..5517	CDS	hypothetical protein
LOCUS_00013	sequence1	complement (6059..7228)	CDS	integrase
LOCUS_00014	sequence1	complement (7407..8267)	CDS	hypothetical protein
LOCUS_00015	sequence1	complement (8301..8813)	CDS	hypothetical protein



W10-2：アノテーション

②[5839, 11509 bp]のアノテーションの残り。
③[11510, 37328 bp]のアノテーションの一部

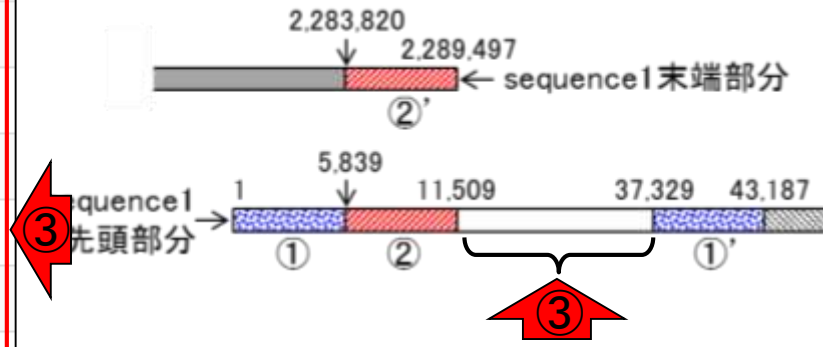
LOCUS_00016	sequence1	complement (8911..9315)	CDS	hypothetical protein
LOCUS_00017	sequence1	complement (9317..9640)	CDS	XRE family transcriptional regulator
LOCUS_00018	sequence1	9916..10098	CDS	hypothetical protein
LOCUS_00019	sequence1	10104..10265	CDS	hypothetical protein
LOCUS_00020	sequence1	10317..11030	CDS	phage-related antirepressor
LOCUS_00021	sequence1	11068..11355	CDS	hypothetical protein
LOCUS_00022	sequence1	11395..11490	CDS	hypothetical protein
LOCUS_00023	sequence1	11503..11736	CDS	hypothetical protein
LOCUS_00024	sequence1	complement (11714..12211)	CDS	hypothetical protein
LOCUS_00025	sequence1	12396..12629	CDS	hypothetical protein
LOCUS_00026	sequence1	12616..13254	CDS	hypothetical protein
LOCUS_00027	sequence1	13258..13974	CDS	phage NTP-binding protein
LOCUS_00028	sequence1	13977..14576	CDS	hypothetical protein
LOCUS_00029	sequence1	14593..14688	CDS	hypothetical protein
LOCUS_00030	sequence1	14689..15522	CDS	hypothetical protein



W10-2：アノテーション

③の領域②[11510, 37328 bp]
のアノテーションの一部

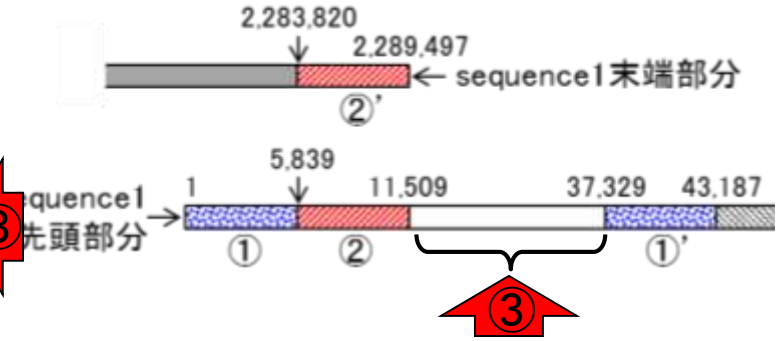
LOCUS_00031	sequence1	15519..16757	CDS	replicative DNA helicase
LOCUS_00032	sequence1	16760..17158	CDS	hypothetical protein
LOCUS_00033	sequence1	17162..17302	CDS	hypothetical protein
LOCUS_00034	sequence1	17306..17671	CDS	hypothetical protein
LOCUS_00035	sequence1	17668..17859	CDS	hypothetical protein
LOCUS_00036	sequence1	17908..18282	CDS	hypothetical protein
LOCUS_00037	sequence1	18279..18482	CDS	hypothetical protein
LOCUS_00038	sequence1	18533..18700	CDS	hypothetical protein
LOCUS_00039	sequence1	18684..18857	CDS	hypothetical protein
LOCUS_00040	sequence1	18859..19314	CDS	Holliday junction resolvase
LOCUS_00041	sequence1	19314..19559	CDS	hypothetical protein
LOCUS_00042	sequence1	19649..20095	CDS	phage protein
LOCUS_00043	sequence1	20310..20585	CDS	hypothetical protein
LOCUS_00044	sequence1	20578..21057	CDS	hypothetical protein
LOCUS_00045	sequence1	21180..21698	CDS	phage terminase small subunit



W10-2：アノテーション

③の領域②[11510, 37328 bp]のアノテーションの残り

LOCUS_00046	sequence1	21698..23599	CDS	phage terminase large subunit
LOCUS_00047	sequence1	23609..23785	CDS	hypothetical protein
LOCUS_00048	sequence1	23786..24961	CDS	phage portal protein
LOCUS_00049	sequence1	24942..26834	CDS	phage capsid protein
LOCUS_00050	sequence1	27026..27313	CDS	hypothetical protein
LOCUS_00051	sequence1	27294..27671	CDS	hypothetical protein
LOCUS_00052	sequence1	27674..28096	CDS	hypothetical protein
LOCUS_00053	sequence1	28096..28476	CDS	hypothetical protein
LOCUS_00054	sequence1	28480..29088	CDS	phage tail protein
LOCUS_00055	sequence1	29239..29589	CDS	tail protein
LOCUS_00056	sequence1	29793..34418	CDS	phage tail tape measure protein
LOCUS_00057	sequence1	34434..35273	CDS	phage protein
LOCUS_00058	sequence1	35230..36441	CDS	prophage protein
LOCUS_00059	sequence1	36425..36751	CDS	hypothetical protein
LOCUS_00060	sequence1	36751..38214	CDS	hypothetical protein

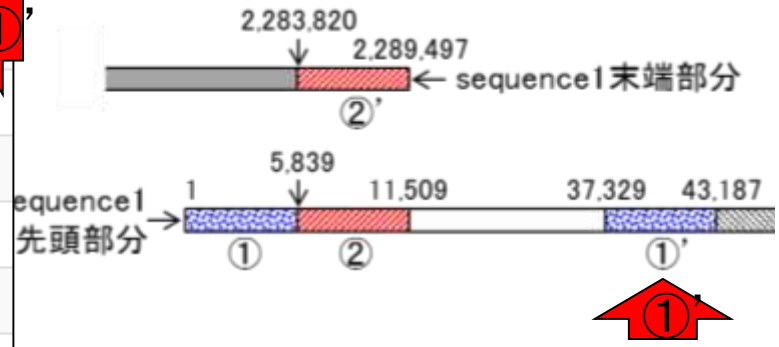


①’と①のアノテーション結果は確かに同じ

W10-2：アノテーション

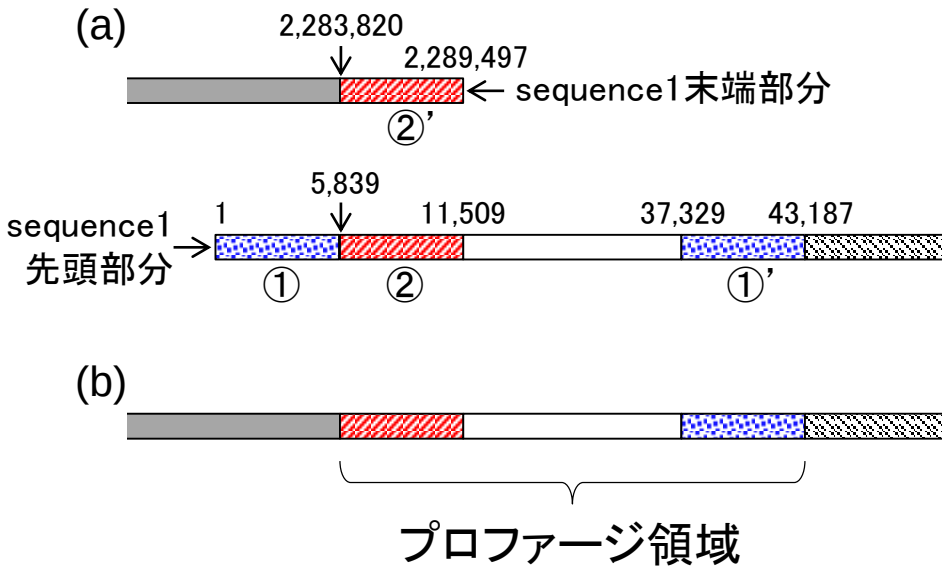
LOCUS_00061	sequence1	38211..38966	CDS	hypothetical protein
LOCUS_00062	sequence1	38966..39298	CDS	hypothetical protein
LOCUS_00063	sequence1	39295..39492	CDS	hypothetical protein
LOCUS_00064	sequence1	39683..40060	CDS	prophage protein
LOCUS_00065	sequence1	40053..40268	CDS	hypothetical protein
LOCUS_00066	sequence1	40394..40861	CDS	holin
LOCUS_00067	sequence1	40874..41905	CDS	1,4-beta-N-acetylmuramidase
LOCUS_00068	sequence1	42092..42262	CDS	hypothetical protein
LOCUS_00069	sequence1	42263..42844	CDS	hypothetical protein
LOCUS_00070	sequence1	43586..44437	CDS	homoserine O-succinyltransferase
LOCUS_00071	sequence1	44450..45724	CDS	O-acetylhomoserine aminocarboxypropyltransferase
LOCUS_00072	sequence1	45808..46845	CDS	competence protein CoiA
LOCUS_00073	sequence1	complement (46909..47541)	CDS	dithiol-disulfide isomerase
LOCUS_00074	sequence1	47818..48507	CDS	GTP pyrophosphokinase
LOCUS_00075	sequence1	48519..49328	CDS	ATP-NAD kinase

①’



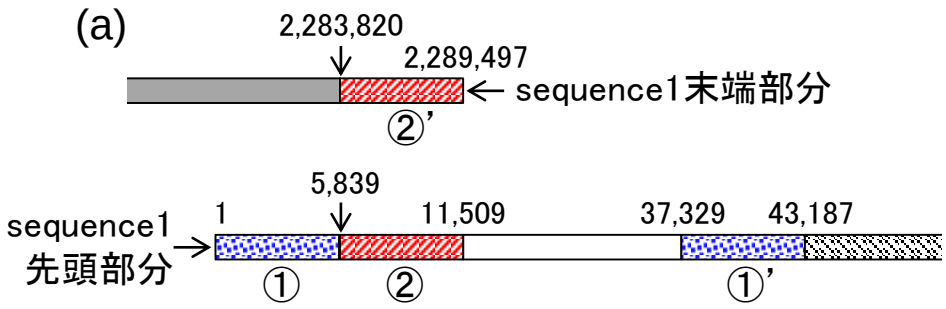
W10-3：染色体構造

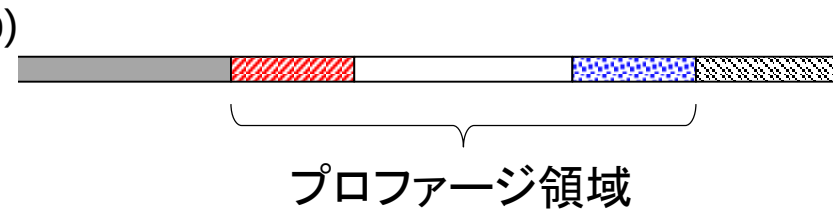
(b) sequence1の両末端である①と②’の領域をトリム。得られた領域[5839, 2283819 bp]の両末端を結合した環状コンティグが実際の染色体構造であると予想した



W10-4：ファージの機構

(b) → (c)プロファージ領域が染色体から切り出されて環状化した状態(excision)





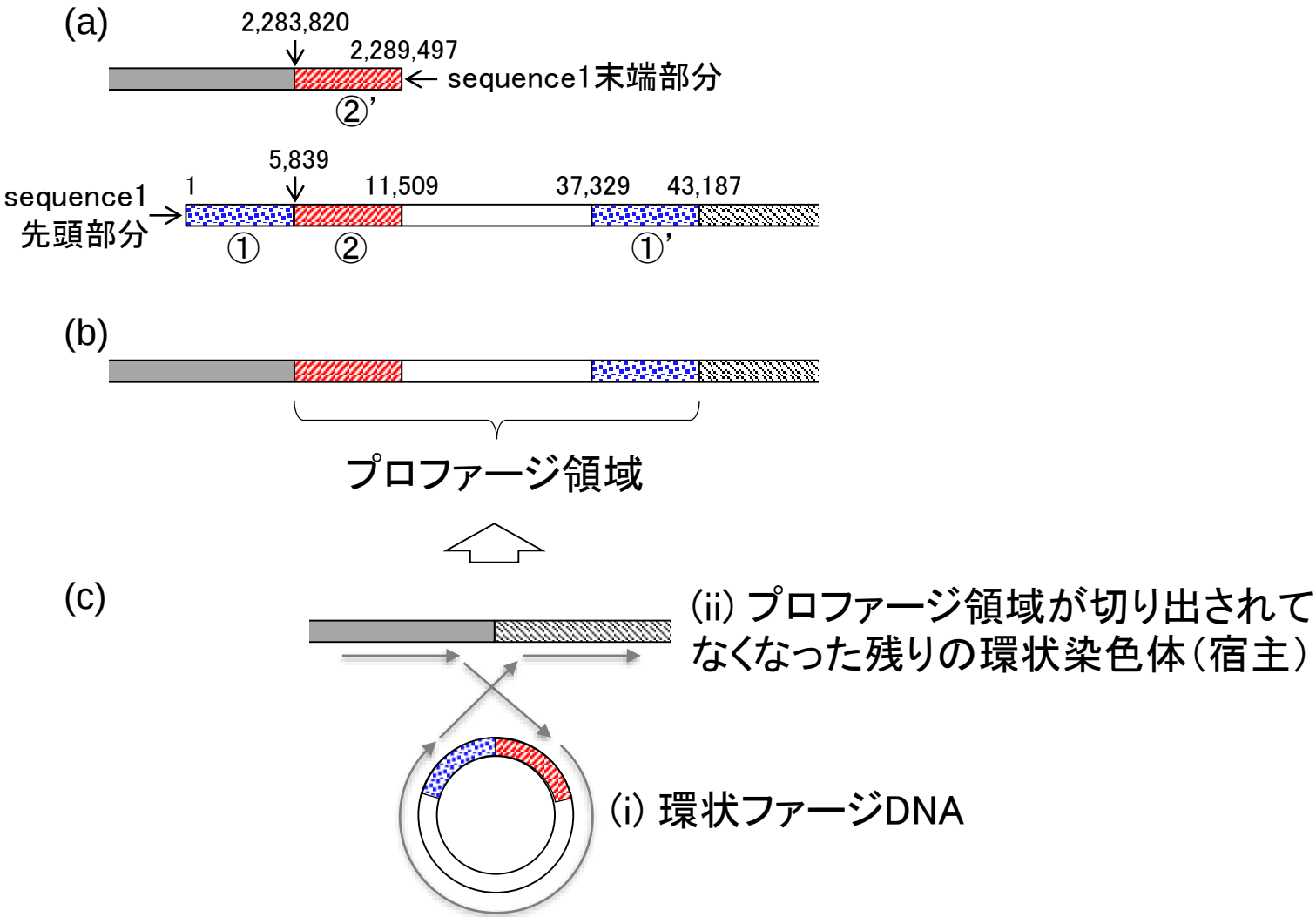
(c)

(ii) プロファージ領域が切り出されてなくなった残りの環状染色体（宿主）

(i) 環状ファージDNA

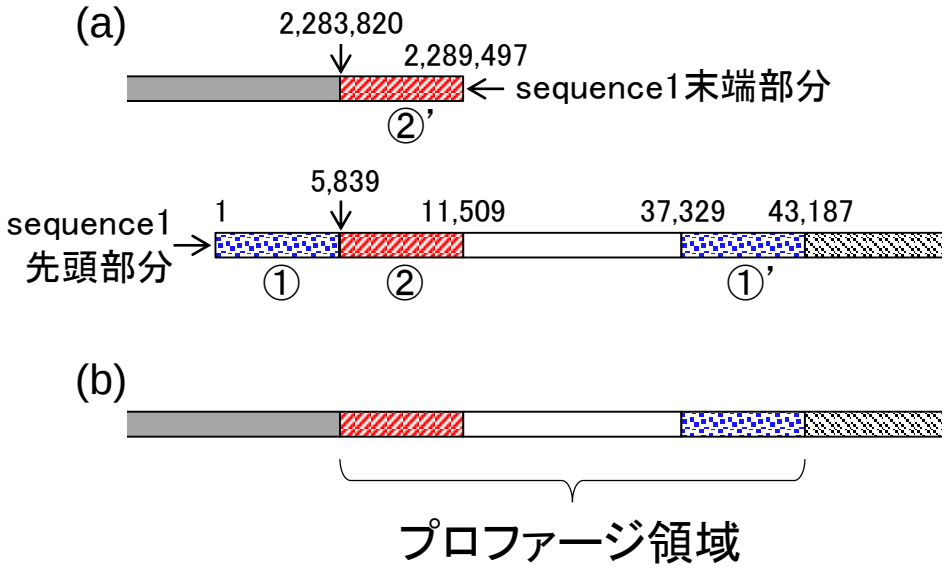
W10-4：ファージの機構

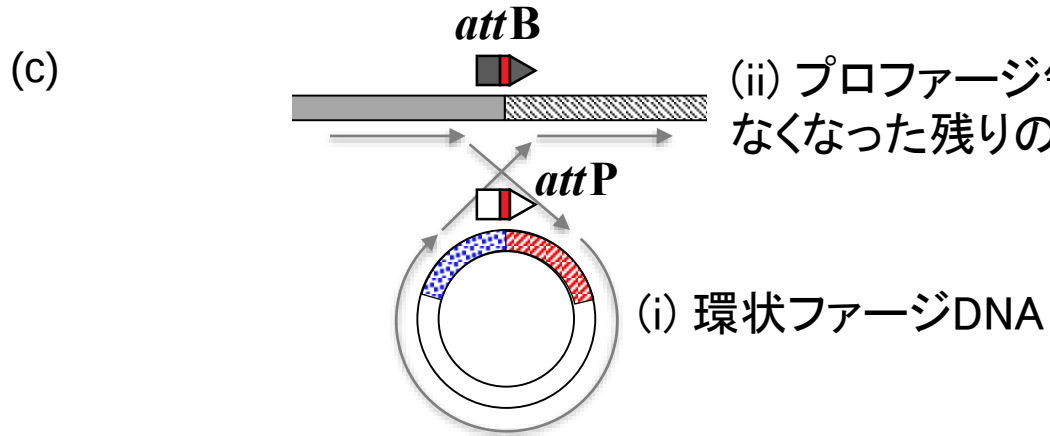
(c) → (b) ファージが染色体に組み込まれる(integration)場合は、灰色矢印の方向に沿って行われる



W10-4：ファージの機構

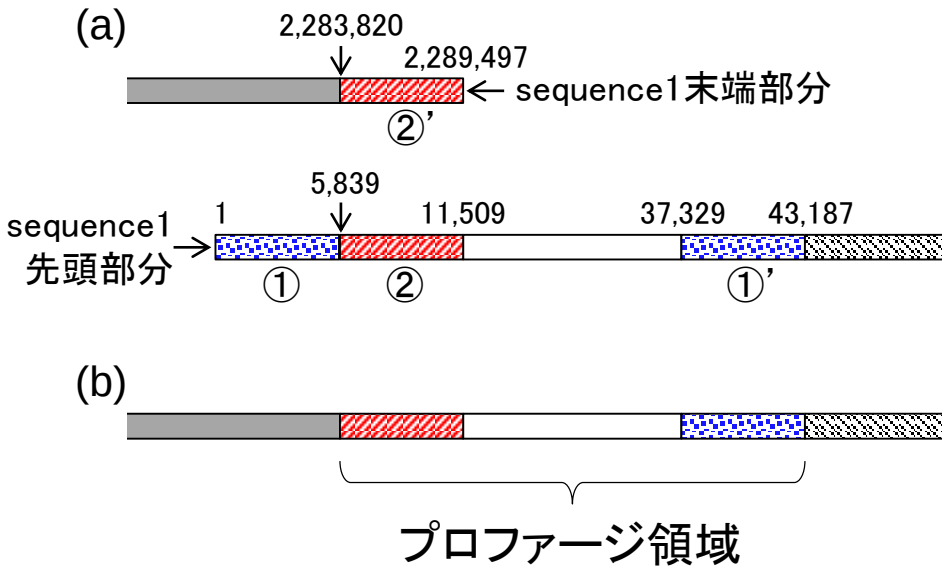
excisionやintegrationが行われるためには、まず染色体上にファージが付着しなければならない。*att*がファージ付着部位(attachment site)であり、赤四角部分が付着部位のコア領域

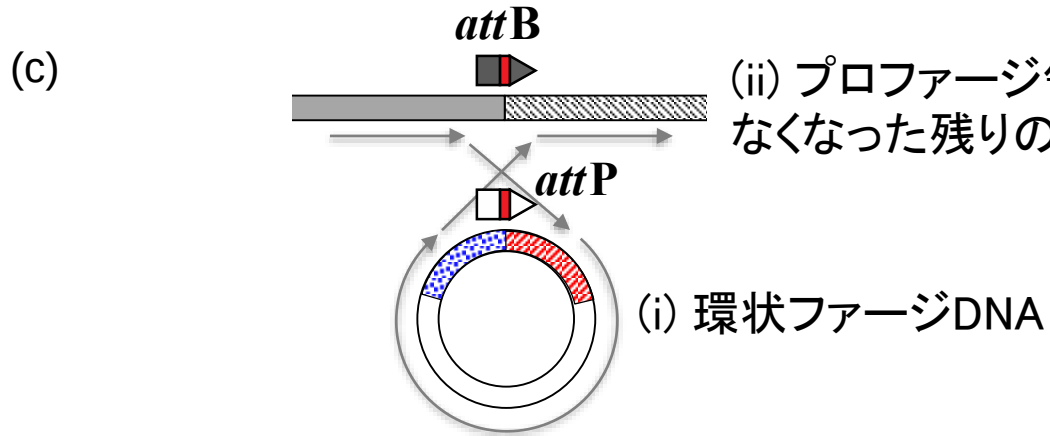




W10-4：ファージの機構

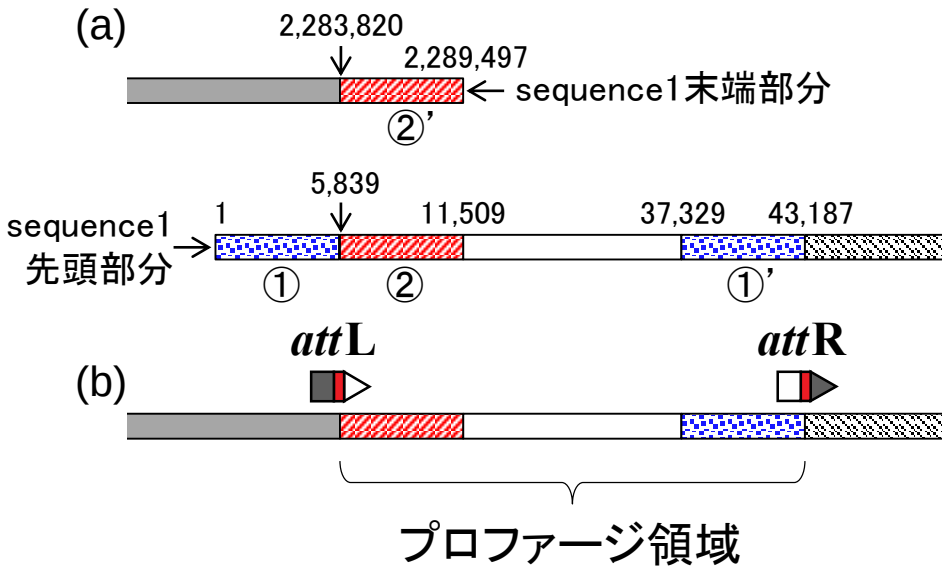
プロファージの特徴1。*att*Bの近傍には、
tRNAがコードされていることが多い（宿主側）

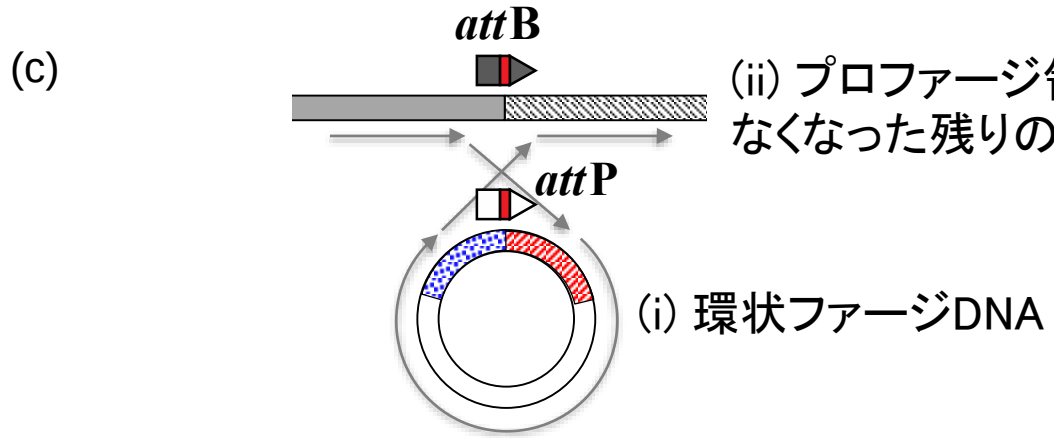




W10-4：ファージの機構

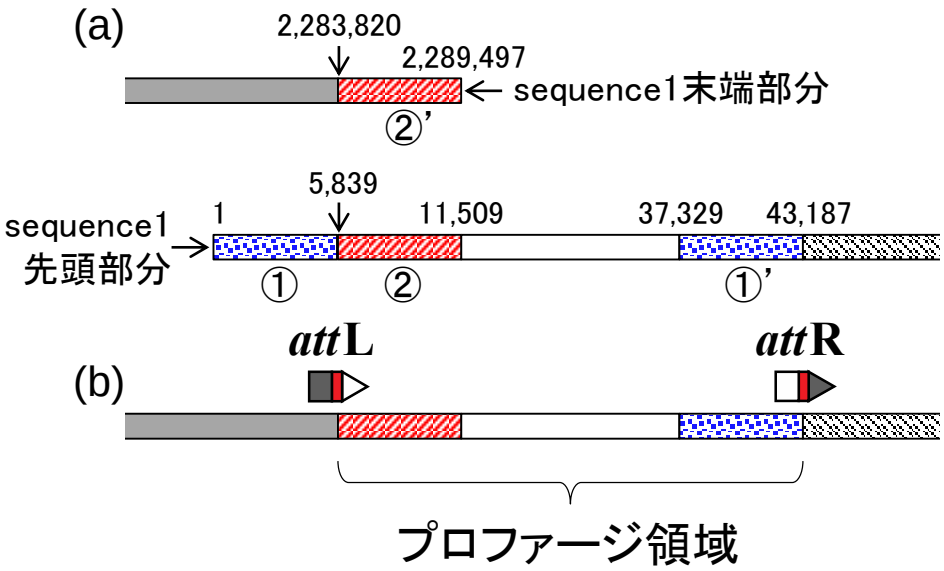
プロファージの特徴2。プロファージ領域の両末端(*att*Lと*att*R)には、10–20塩基の相同な配列が存在する。付着部位のコア領域のことを指し、赤四角部分に相当する。アノテーション情報の特徴としては、一方の末端付近にはintegrase、逆の末端付近にはlysinとholinが存在する

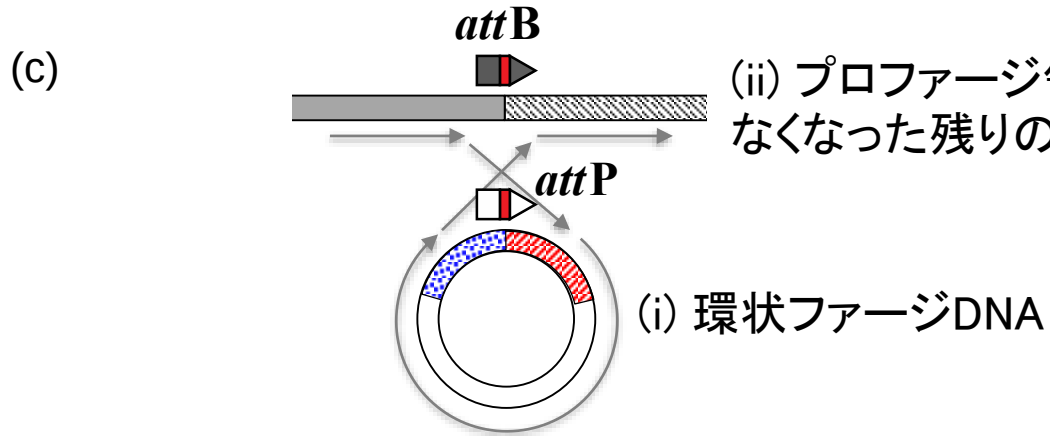




W10-4：ファージの機構

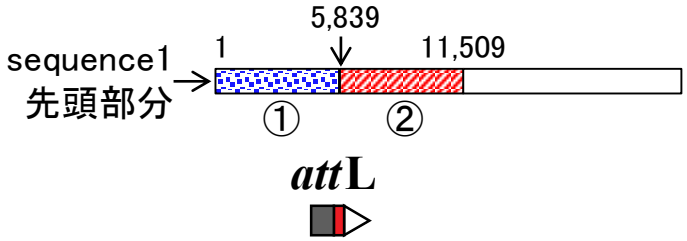
プロファージ領域の右末端(*att*R)は、領域①'の右末端ということ。それは領域①[1, 5860 bp]の右末端と同じ。つまり赤四角部分は、領域①[1, 5860 bp]の右末端と、領域②[5839, 11509 bp]の左末端あたりにあるということです





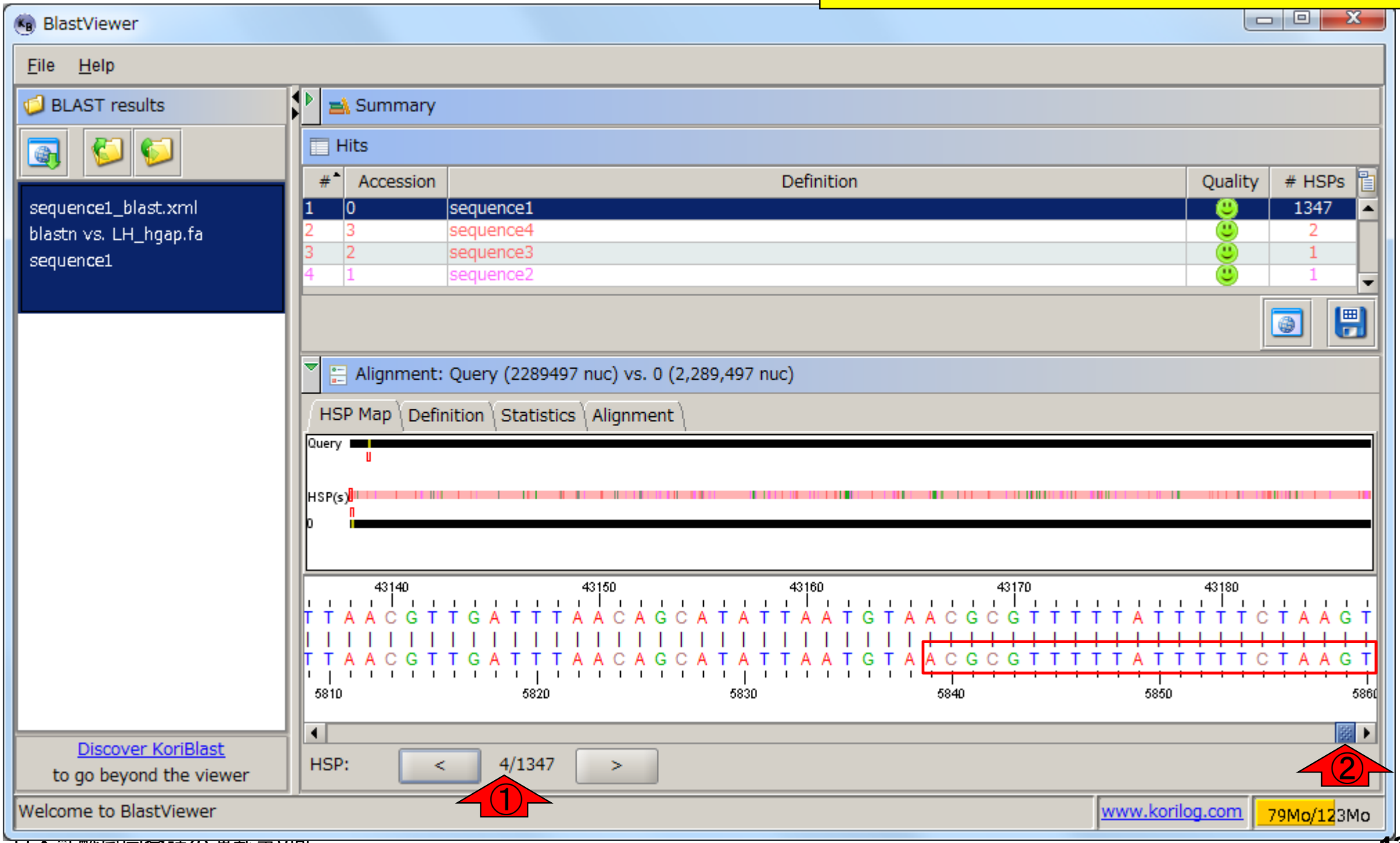
W10-4：ファージの機構

今は左図あたりの議論をしています。領域①[1, 5860 bp]の右末端と、領域②[5839, 11509 bp]の左末端には重複領域[5839, 5860 bp]があることに気づくでしょう。これが赤四角部分に相当



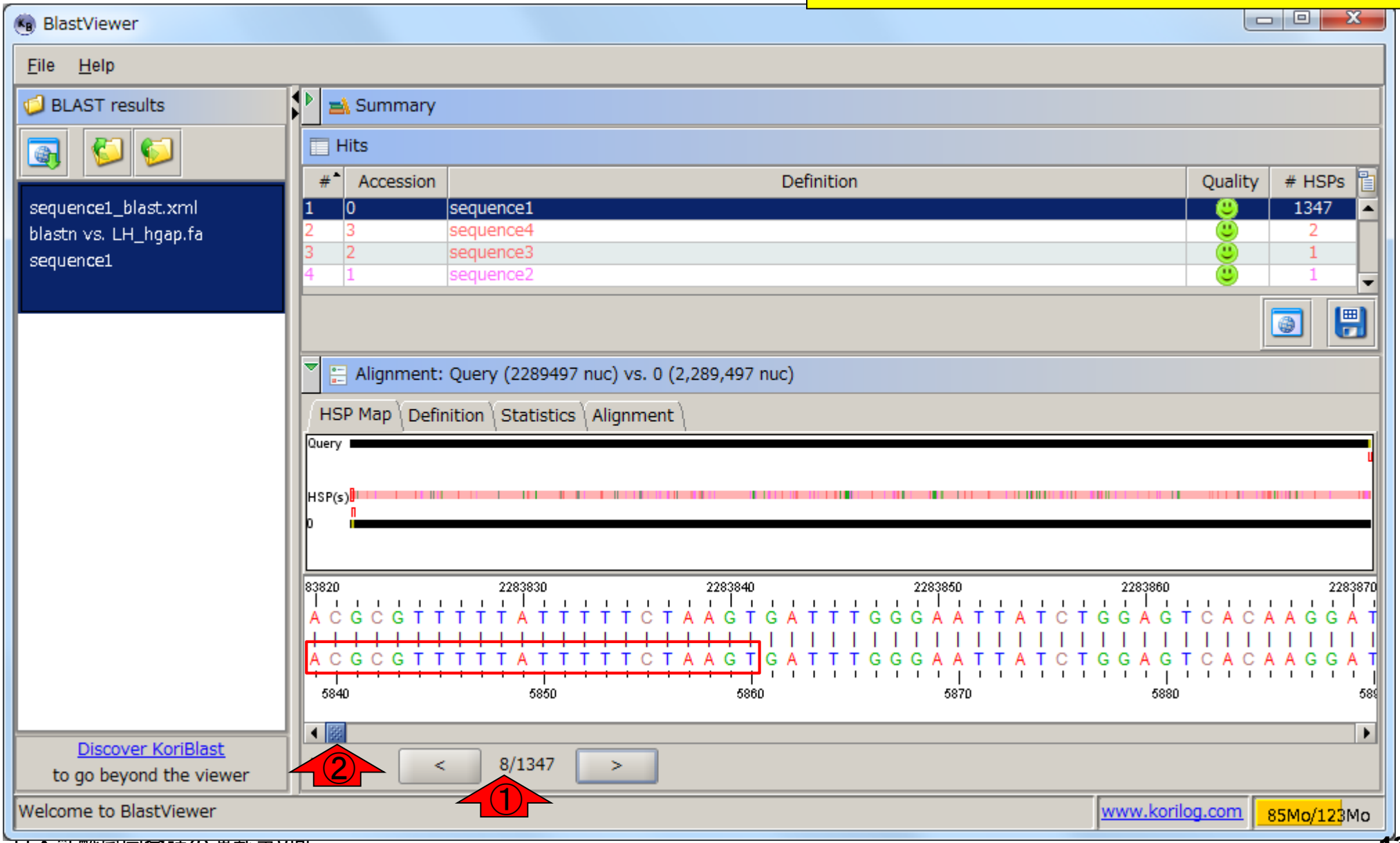
W10-4：ファージの機構

HSP4のアラインメント結果の右端部分を表示。領域①[1, 5860 bp]の右末端に相当。赤枠部分が赤四角部分に相当する



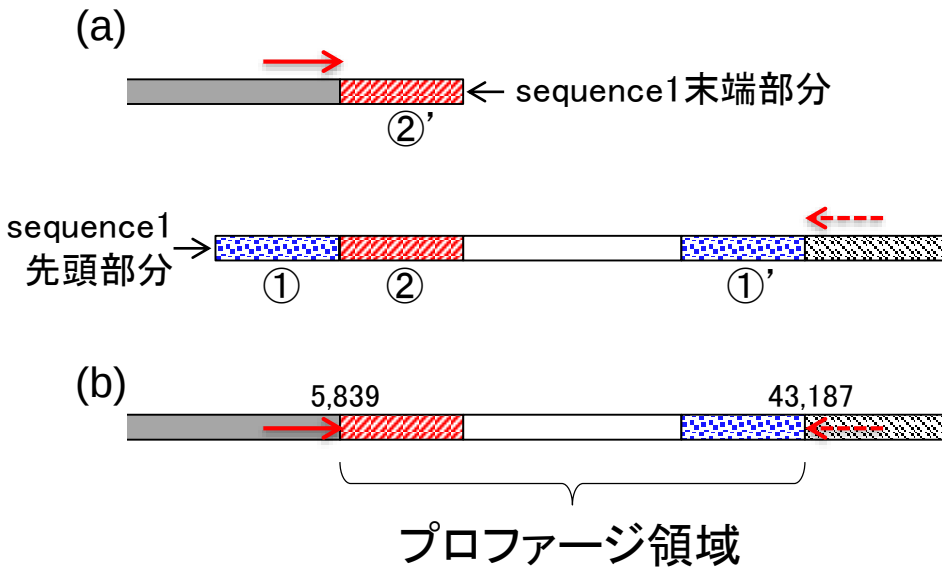
W10-4：ファージの機構

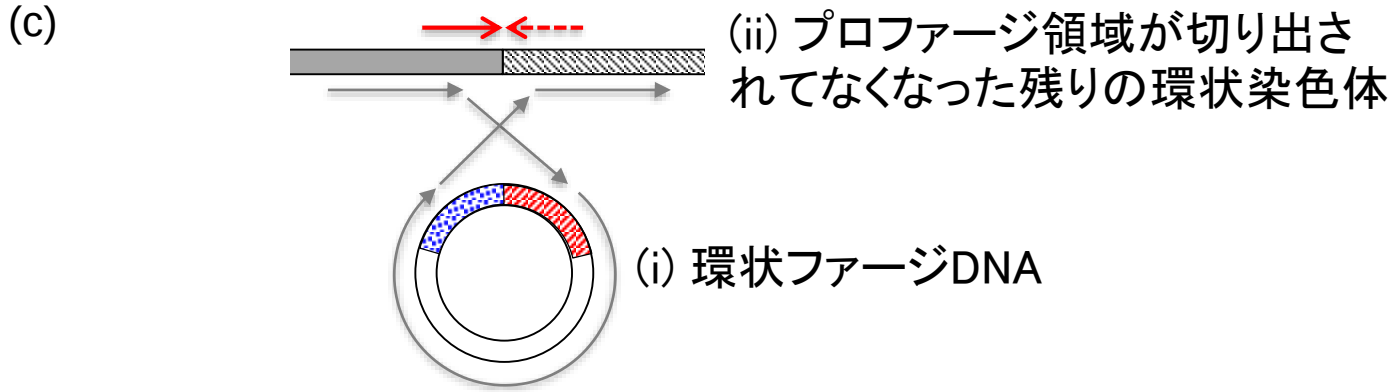
HSP8のアラインメント結果の右端部分を表示。領域②[5839, 11509 bp]の左末端に相当。赤枠部分が赤四角部分に相当する



W10-5：PCR検証1

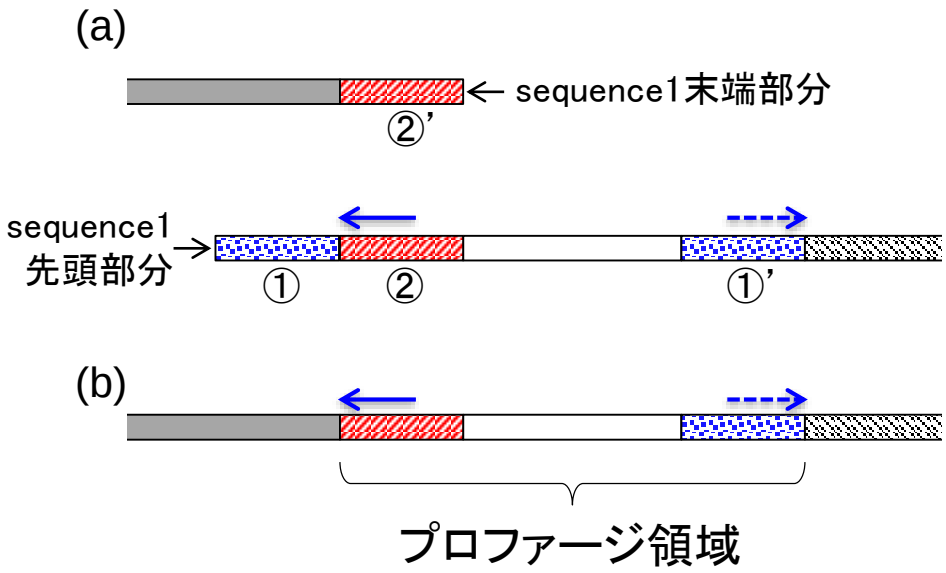
(a) 赤い矢印の位置でPCRのプライマーを設計。(b) プロファージ領域の長さは40kbp程度あるため、通常はPCRで増幅されない。もしシークエンス対象となった細胞集団中に“プロファージ領域を含む環状染色体”しか存在しなければPCR増幅されるものはないはず。(c) シークエンス対象となった細胞集団中には、(b)のプロファージ領域を含む環状染色体以外に、(i) プロファージ領域が切り出されてできた環状ファージDNA、および(ii) プロファージ領域が切り出されてなくなった残りの環状染色体が含まれるとすれば、(ii)の存在比に応じてPCR増幅されるはず。→実際に増幅された





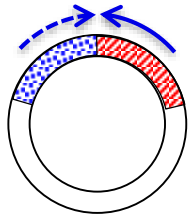
W10-5：PCR検証2

(a) 青い矢印の位置でPCRのプライマーを設計。(b) 逆向きにプライマーを設計しているため、通常はPCRで増幅されない。もしシークエンス対象となった細胞集団中に“プロファージ領域を含む環状染色体”しか存在しなければPCR増幅されるものはないはず。(c) シークエンス対象となった細胞集団中には、(b)のプロファージ領域を含む環状染色体以外に、(i) プロファージ領域が切り出されてできた環状ファージDNA、および(ii) プロファージ領域が切り出されてなくなった残りの環状染色体が含まれるとすれば、(i)の存在比に応じてPCR増幅されるはず。→実際に増幅された



(c)

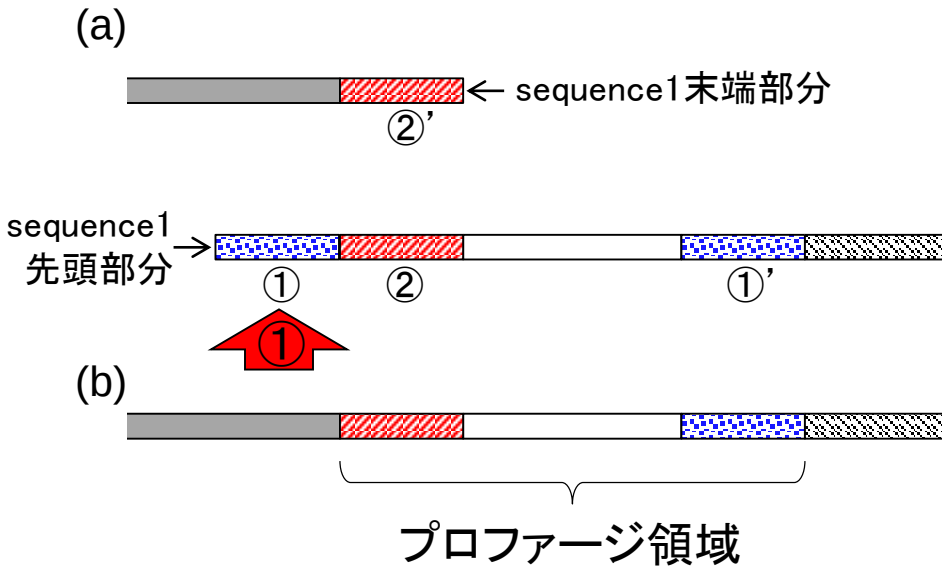
(ii) プロファージ領域が切り出されてなくなった残りの環状染色体



(i) 環状ファージDNA

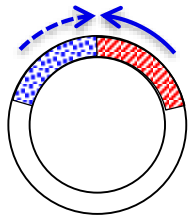
W10-6：①の領域

①の領域は、(i) のように環状化したファージDNAがシークエンスされた結果として生じたものであり、実際の染色体上には存在しないと結論づけた



(c)

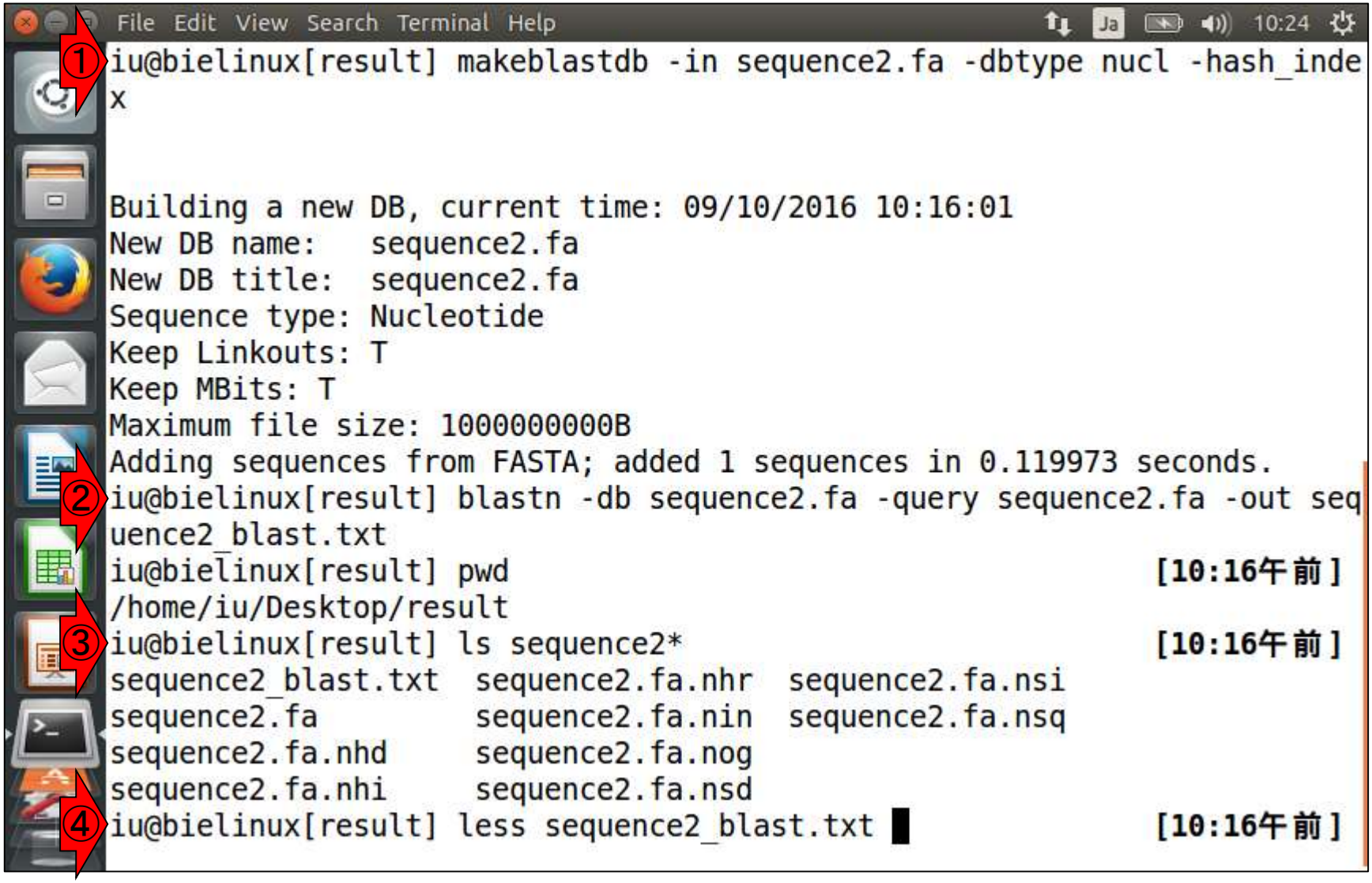
(ii) プロファージ領域が切り出されてなくなった残りの環状染色体



(i) 環状ファージDNA

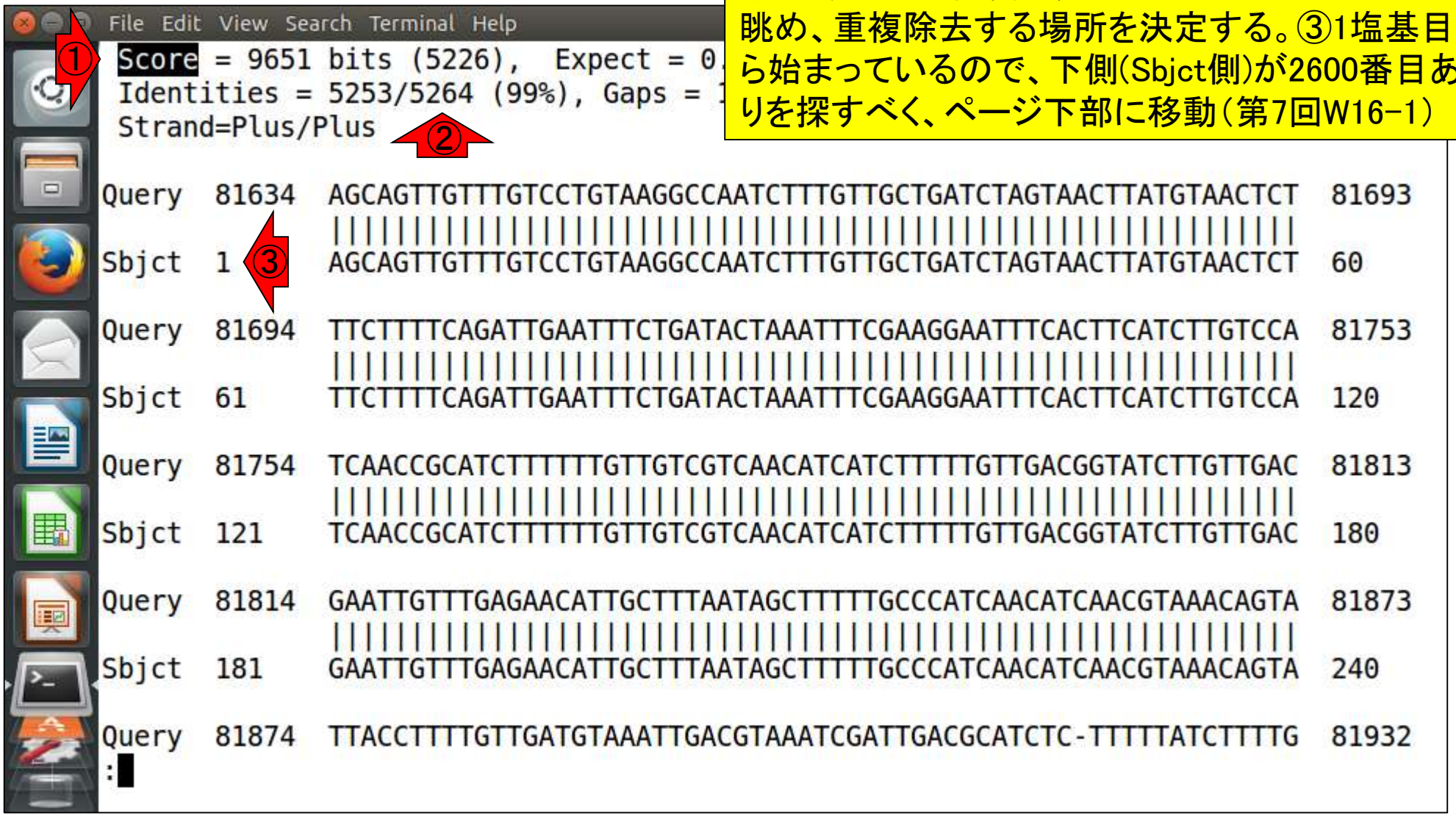
W11-1：BLAST

sequence2のBLASTを実行。①makeblastdb、②blastn。③lsで確認して、④lessで開く



W11-1：BLAST

①Scoreで文字列検索して、2つめのHSPに辿り着いたところ（第7回W15-7）。②長さが5264塩基なので、概ねその半分程度の長さのところのアラインメントを眺め、重複除去する場所を決定する。③1塩基目から始まっているので、下側(Sbjct側)が2600番目あたりを探すべく、ページ下部に移動（第7回W16-1）



W11-1：BLAST

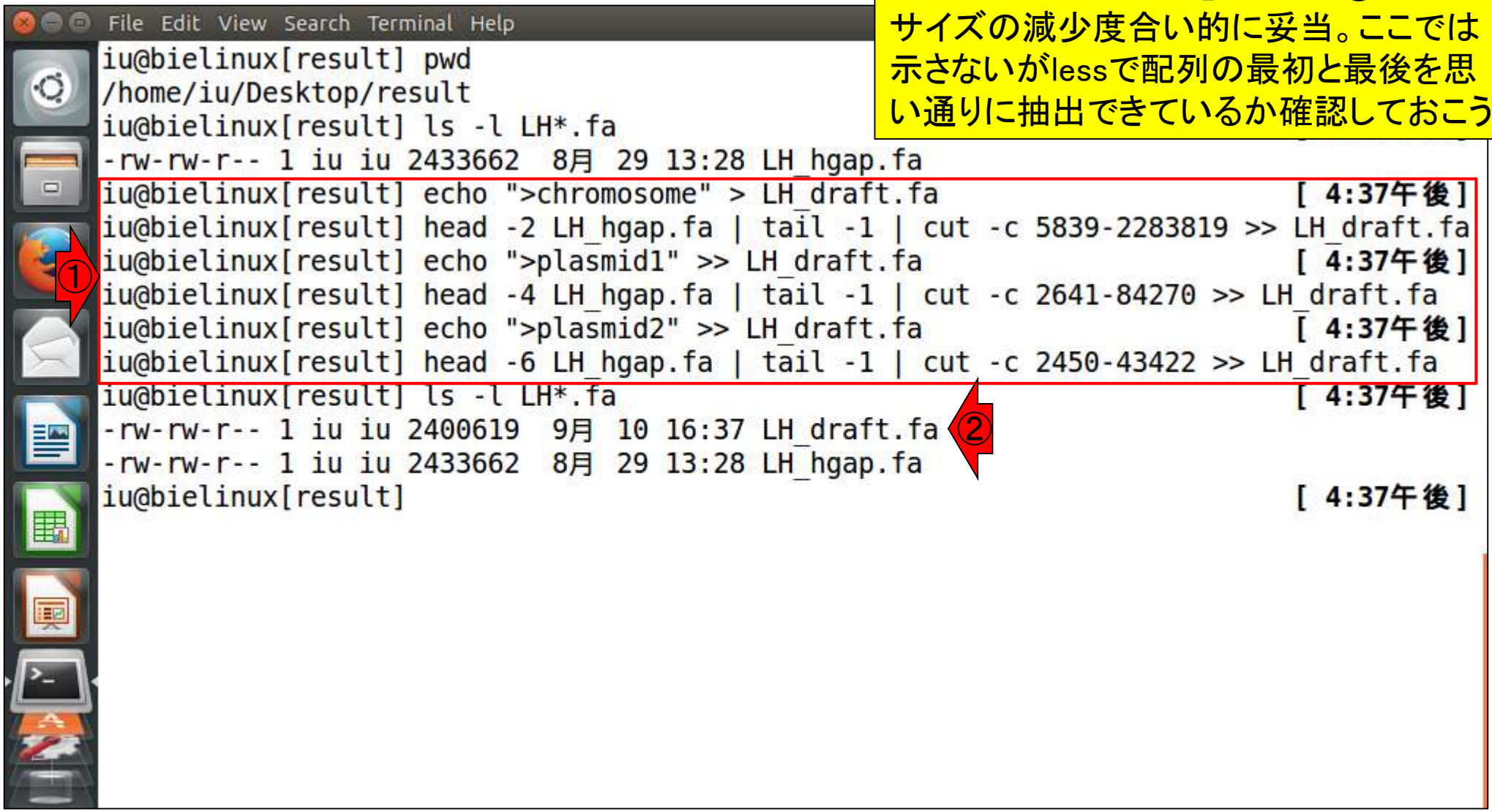
①の赤枠分くらいを眺め、どこにもミスマッチやGapがないことを確認。②のところでトリムすることにする。左端にする理由は、上が84271番目、下が2641番目の塩基だとすぐにわかるから

W11-1：BLAST

①2641番目の塩基をトリム後の1塩基目にする場合は、[2641, 84270 bp]を残せばよい。こうすることで、トリム後の塩基配列の最初のほうは①の赤枠のようになり、最後のほうは②のようになるはずである。③qで終了

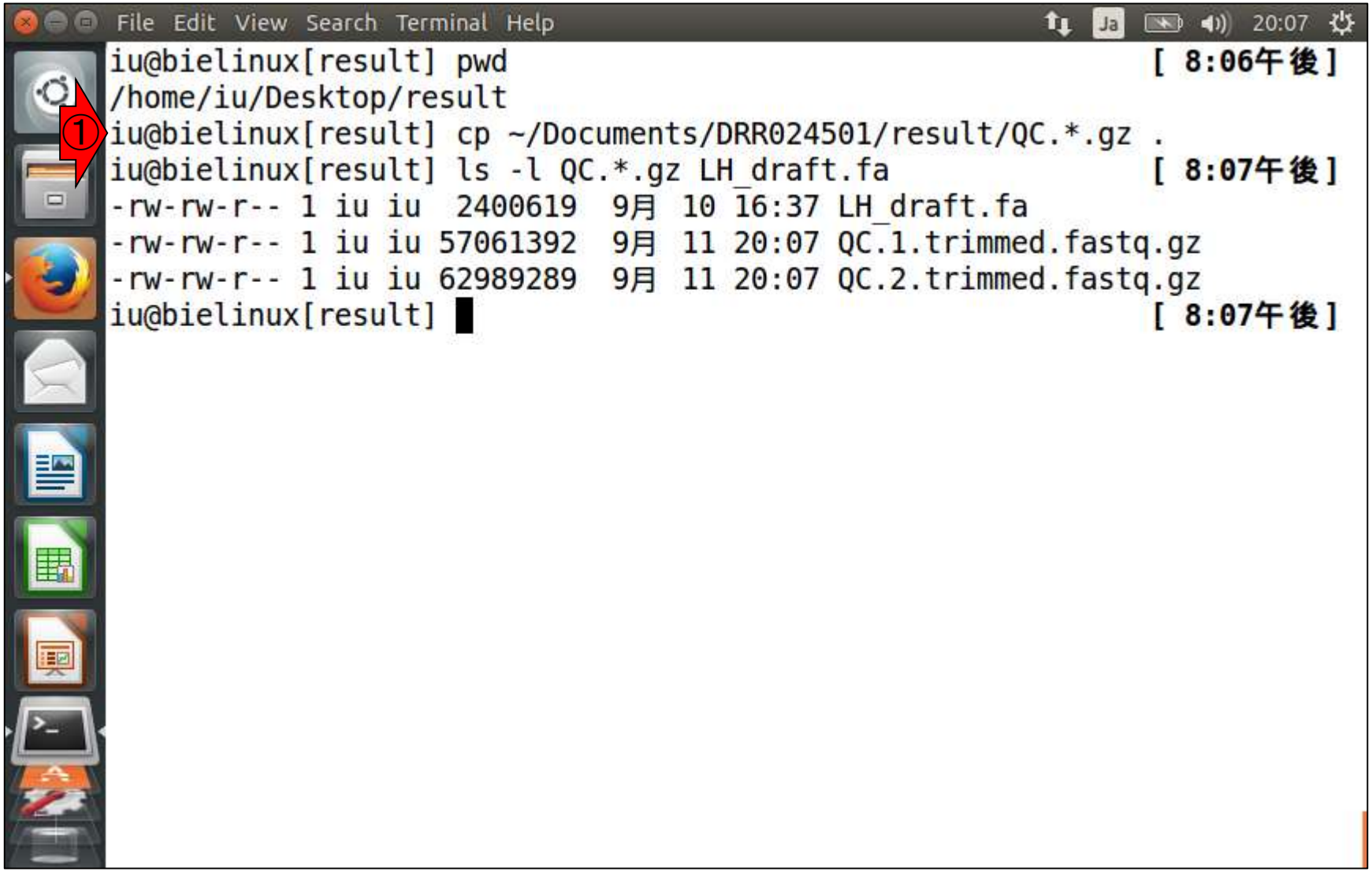
W11-2:ドラフト配列作成

①アセンブリ結果ファイル(LH_hgap.fa)を入力として、重複除去を行う一連のコマンド。出力ファイルはLH_draft.fa。②ファイルサイズの減少度合い的に妥当。ここでは示さないがlessで配列の最初と最後を思い通りに抽出できているか確認しておこう



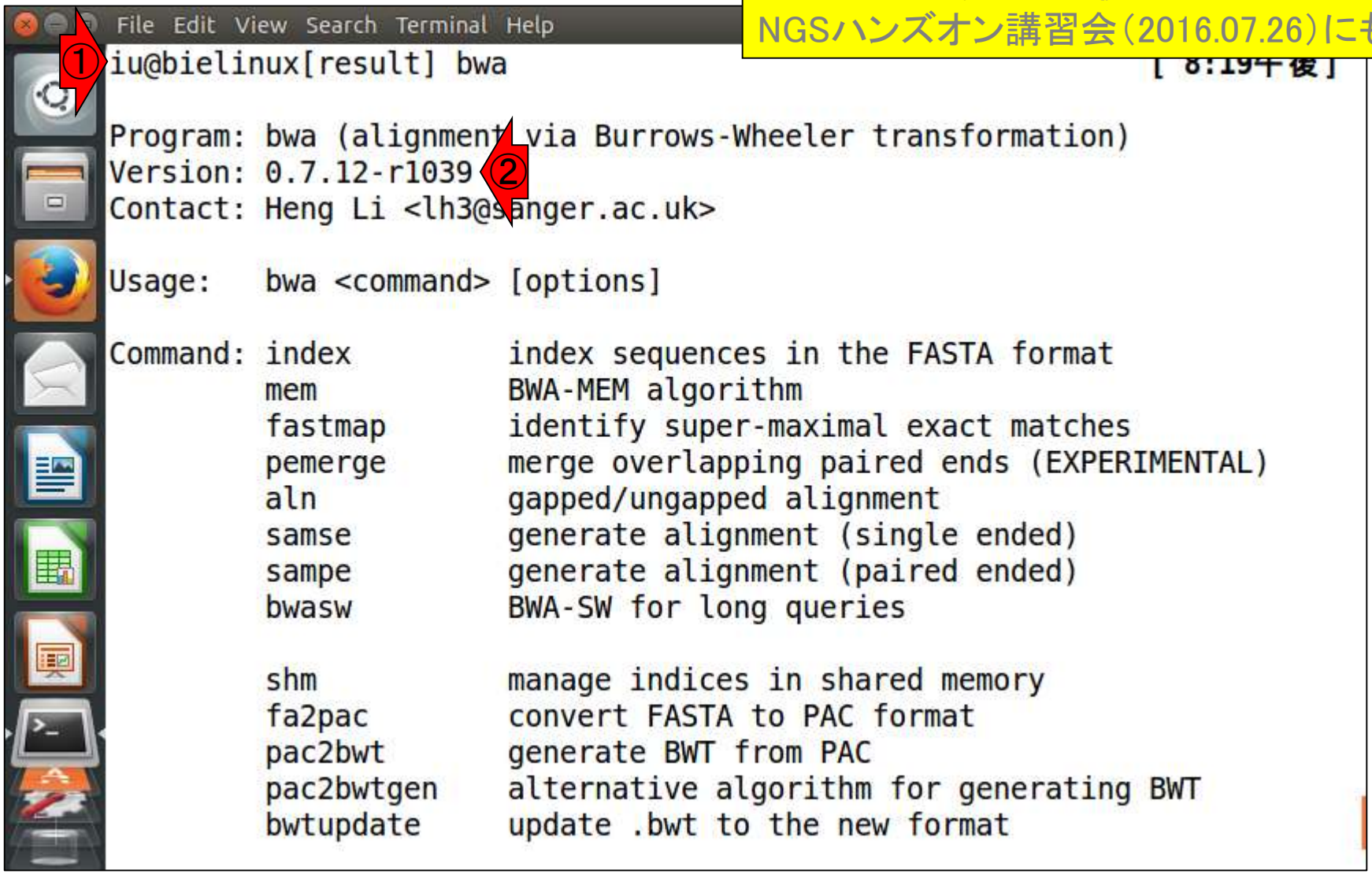
W12-1：マップする側

①~/Documents/DRR024501/result/上にあるQC.*.gzをカレントディレクトリにコピー



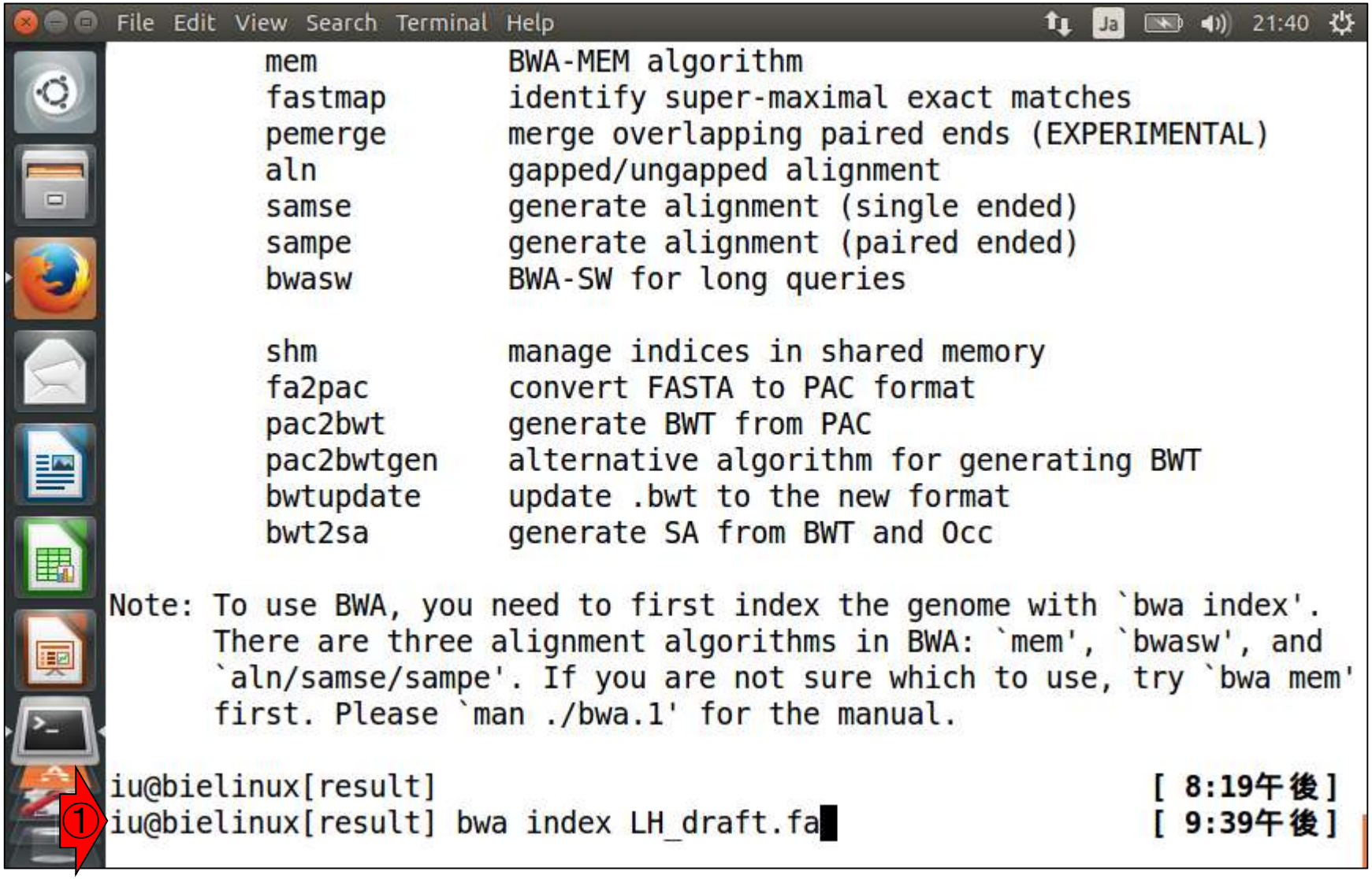
W12-2：bwa

①Bio-Linuxにはbwaがプリインストールされているので、bwaがすぐに利用可能な状態。②バージョンは0.7.12-r1039。BWAの使用法については、平成28年度NGSハンズオン講習会（2016.07.26）にもあり



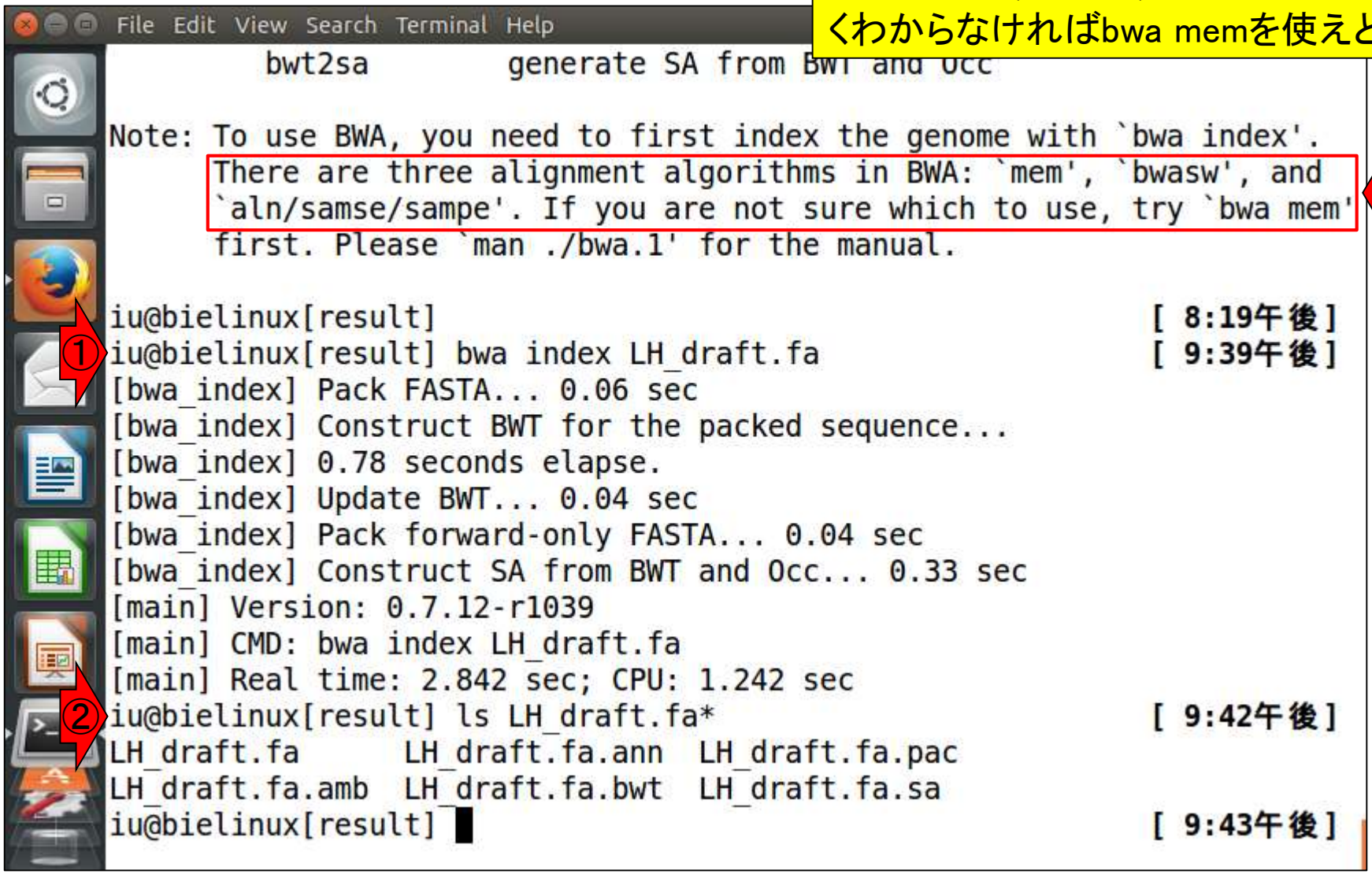
W12-2：bwa

①マップされる側のリファレンス配列(LH_draft.fa)のインデックスファイルを作成



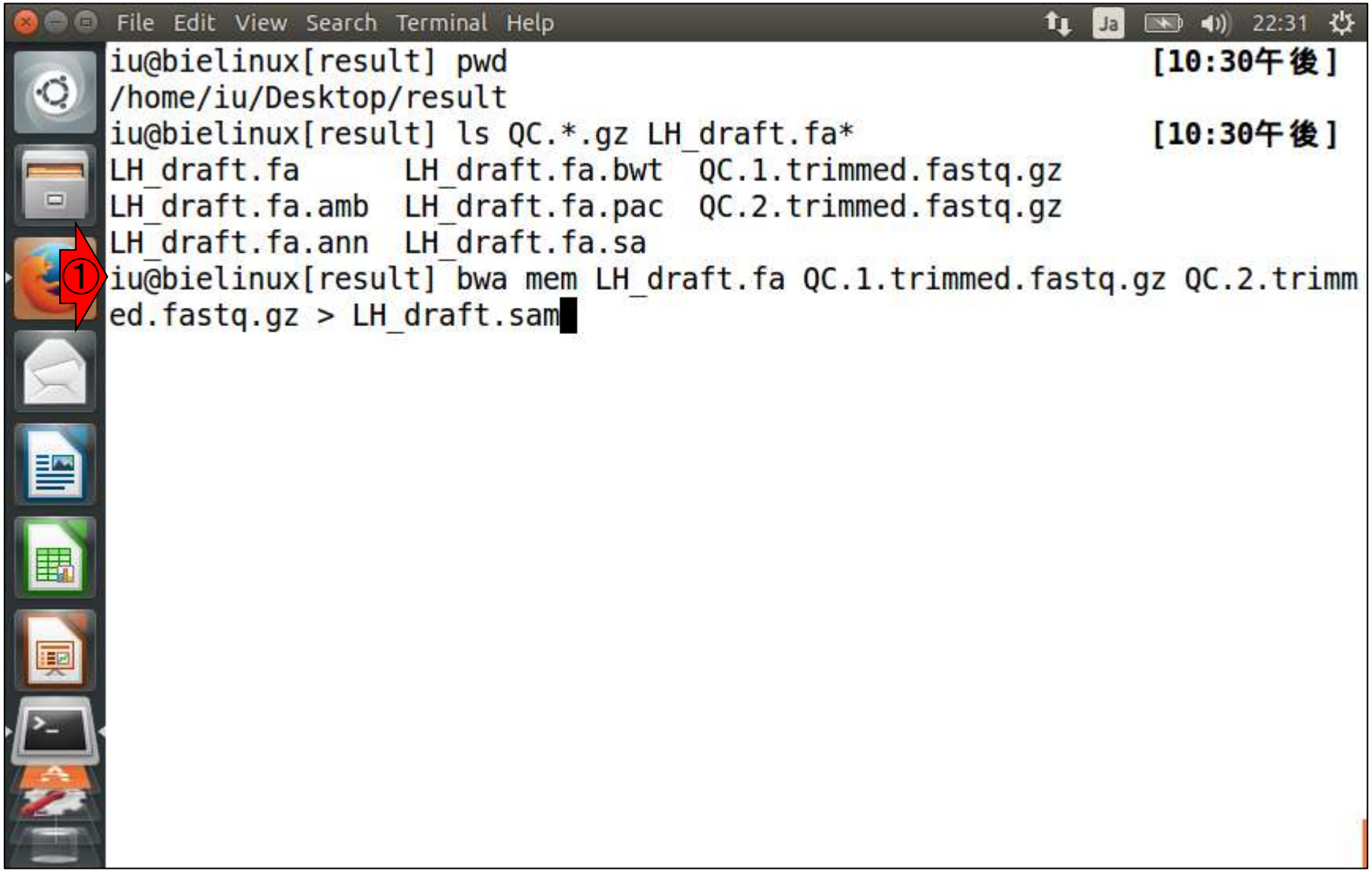
W12-2：bwa

①bwa index実行結果。②5つのインデックスファイルが作成されたようだ。③BWAには3つのアルゴリズム(mem, bwasw, aln/samse/sampe)があり、よくわからなければbwa memを使えと書いている



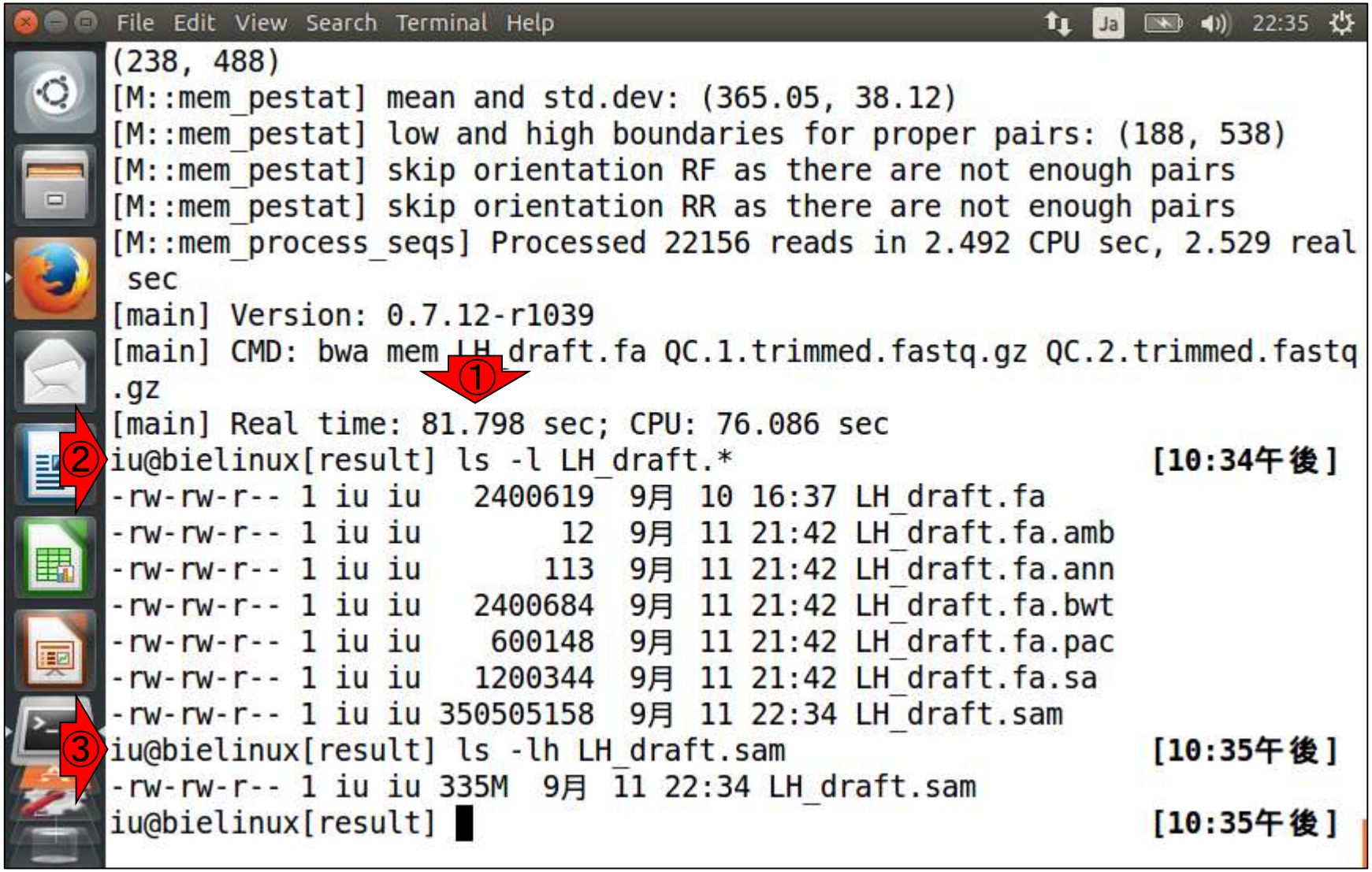
W12-2：bwa

①bwa memの実行。SAM形式のマッピング結果ファイル(LH_draft.sam)を得るところまで。約2分



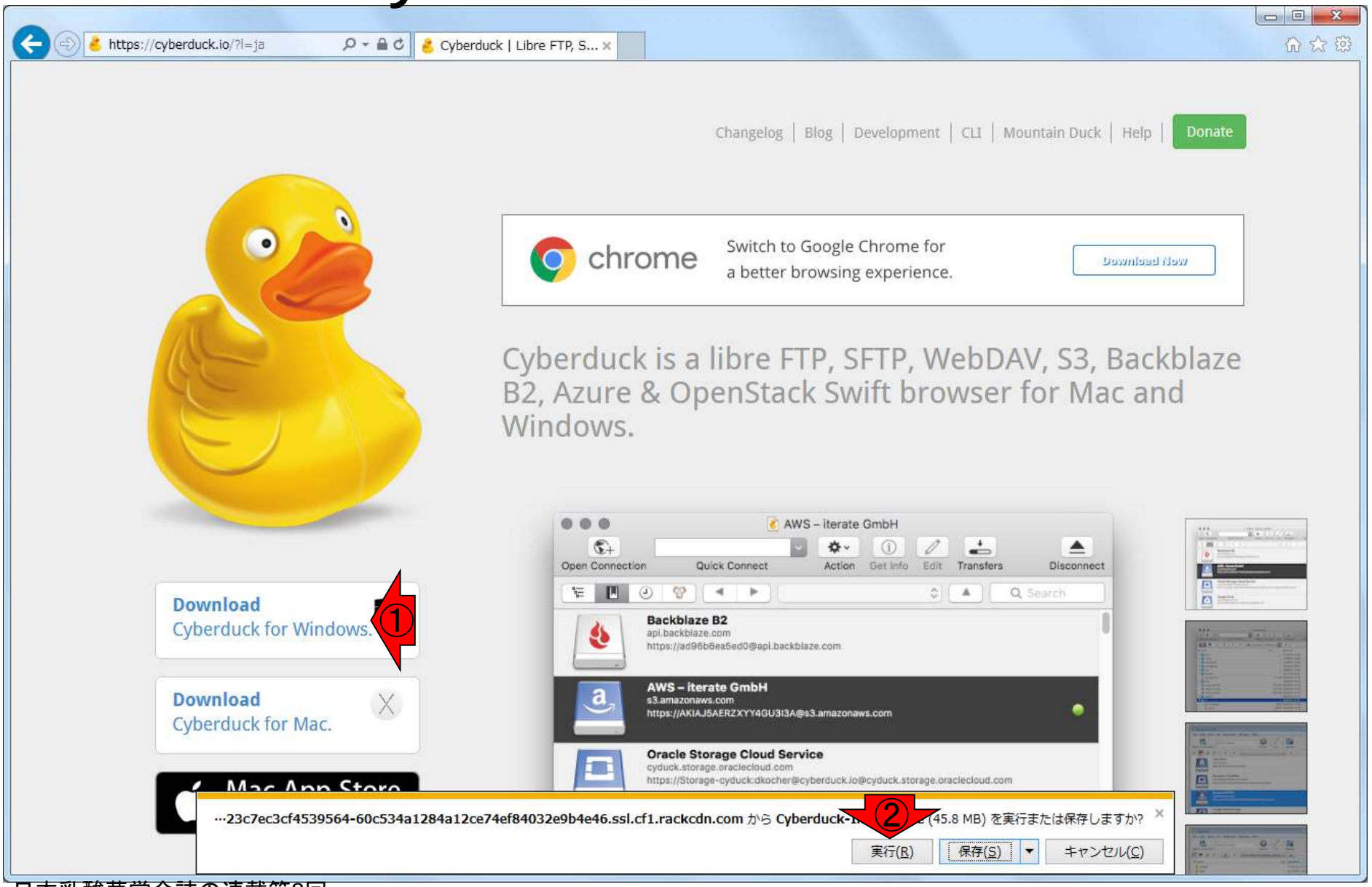
W12-2：bwa

無事計算終了。①実際にかかった時間は約82秒。②lsで確認。③ファイルサイズは約335MB



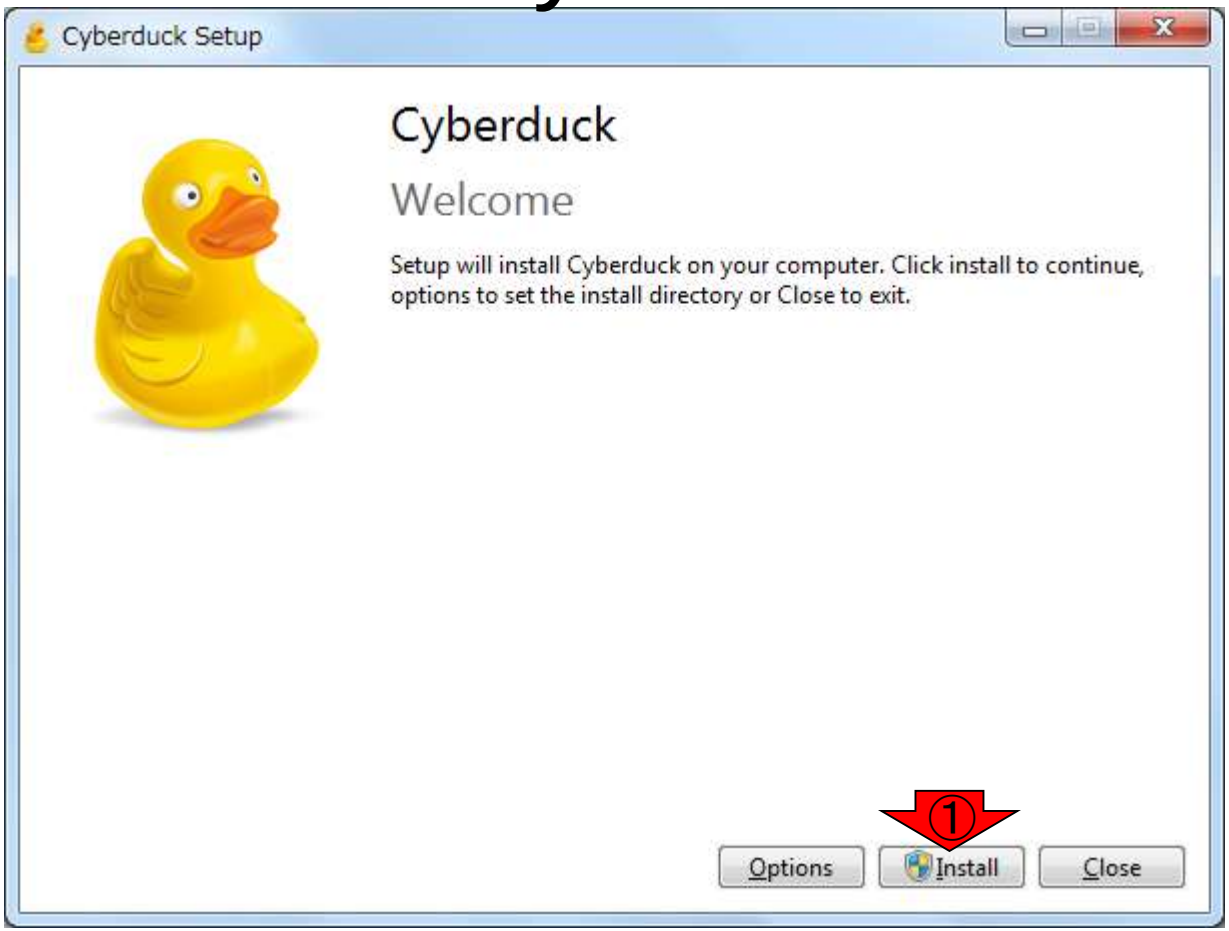
W13-1：Cyberduck

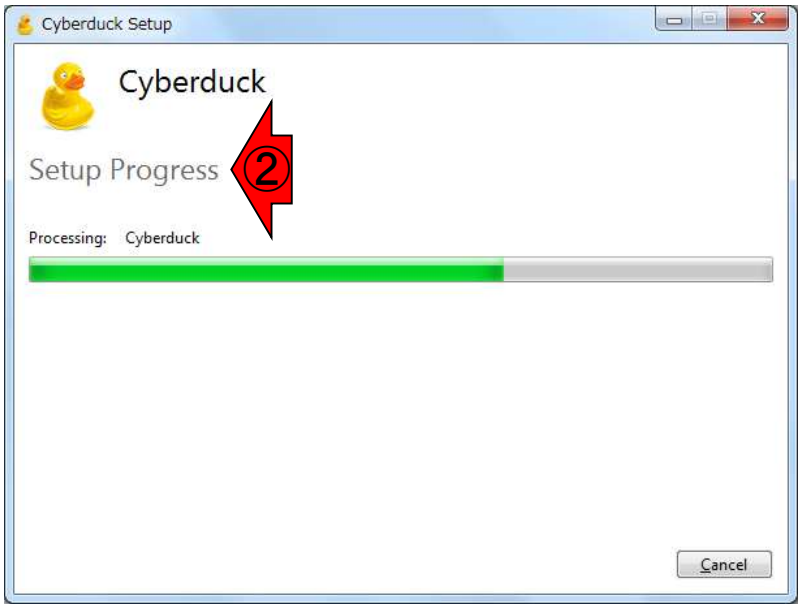
Cyberduckのインストール。①Windows用をインストールしたい場合はこちら。②実行

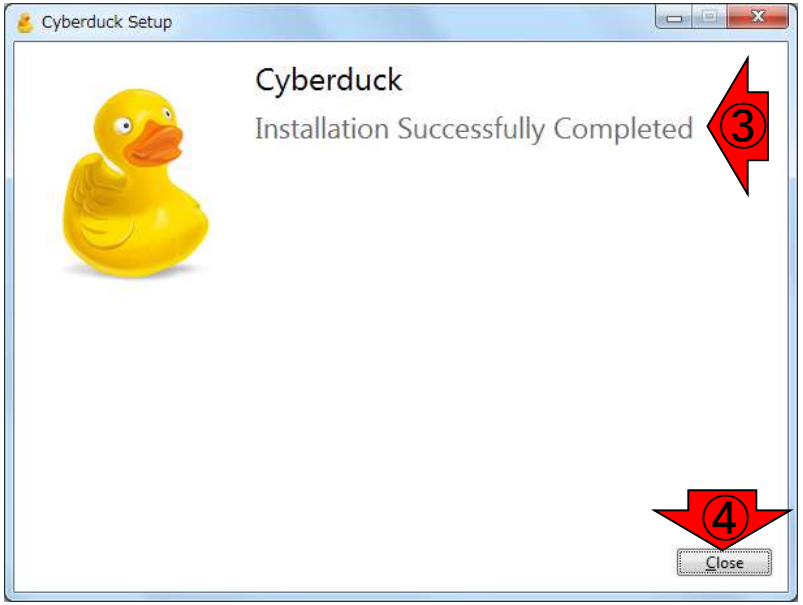


W13-1：Cyberduck

①Install。②途中経過。③無事終了すると右下のような感じになる。④Close

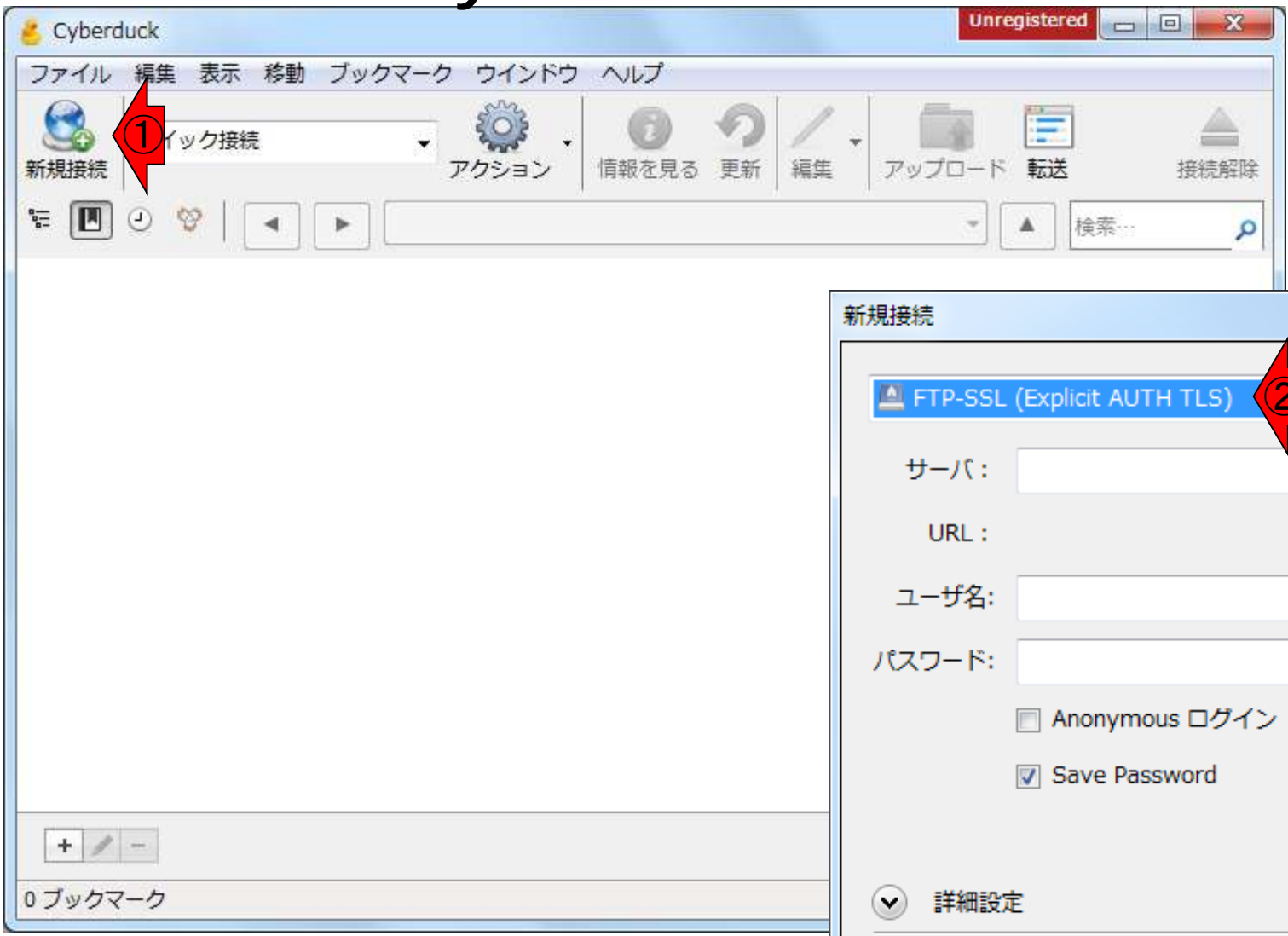


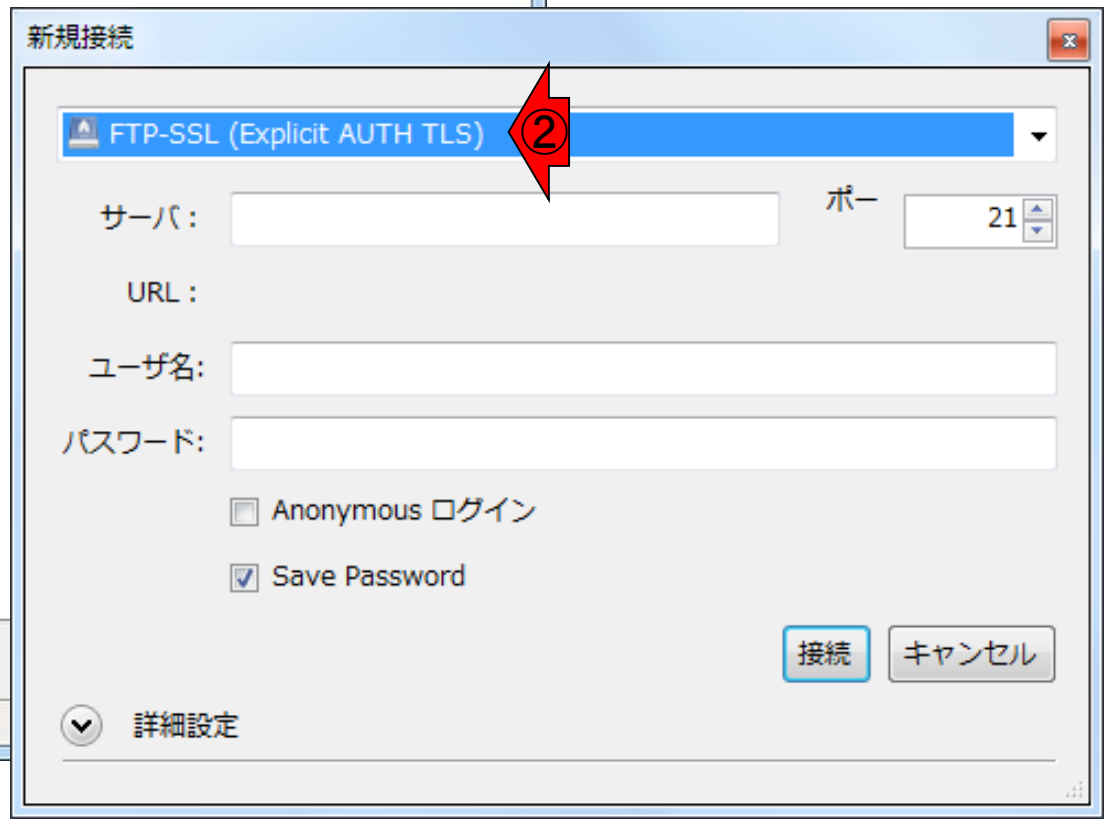




W13-2：Cyberduck設定

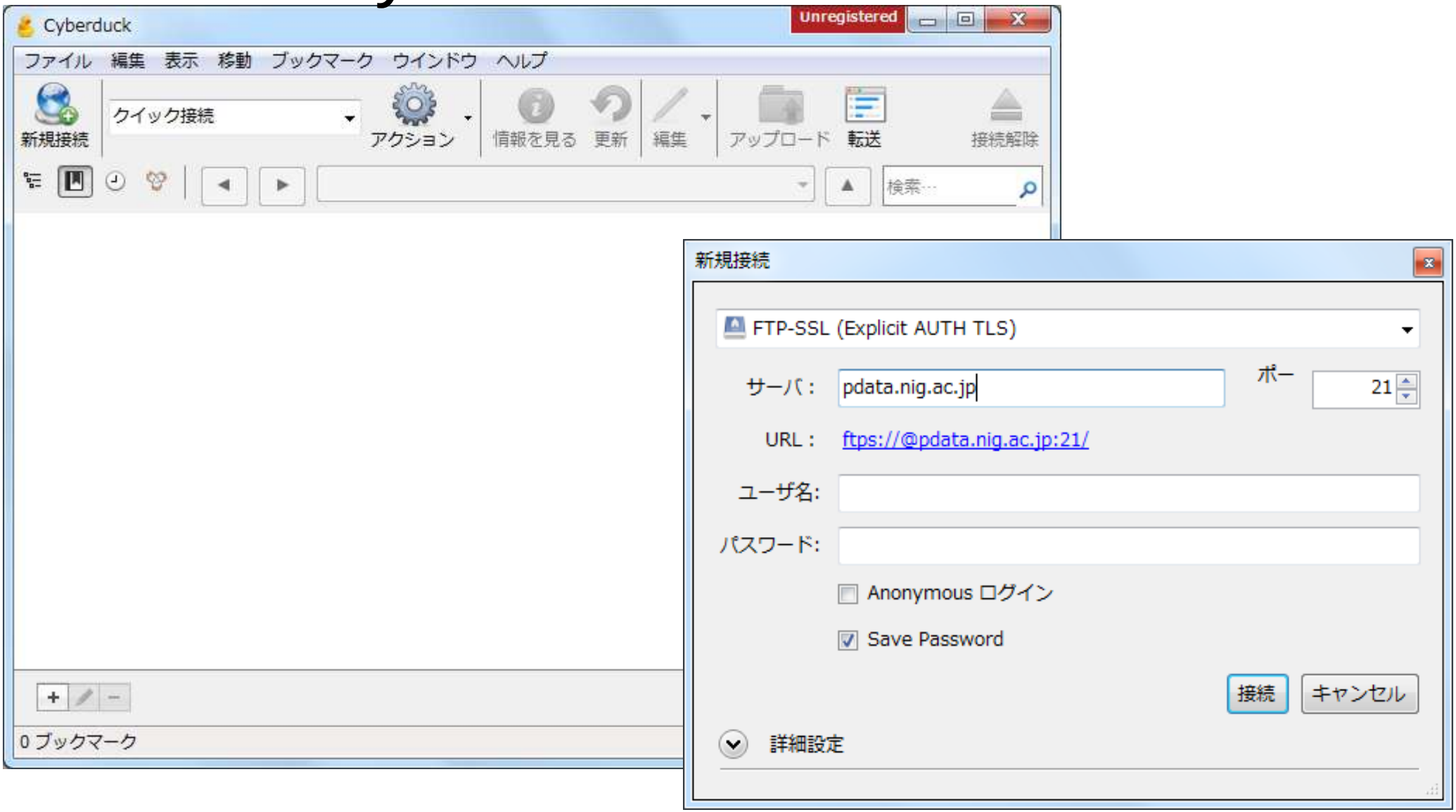
Cyberduckを起動し、①新規接続。②FTP-SSLを選択





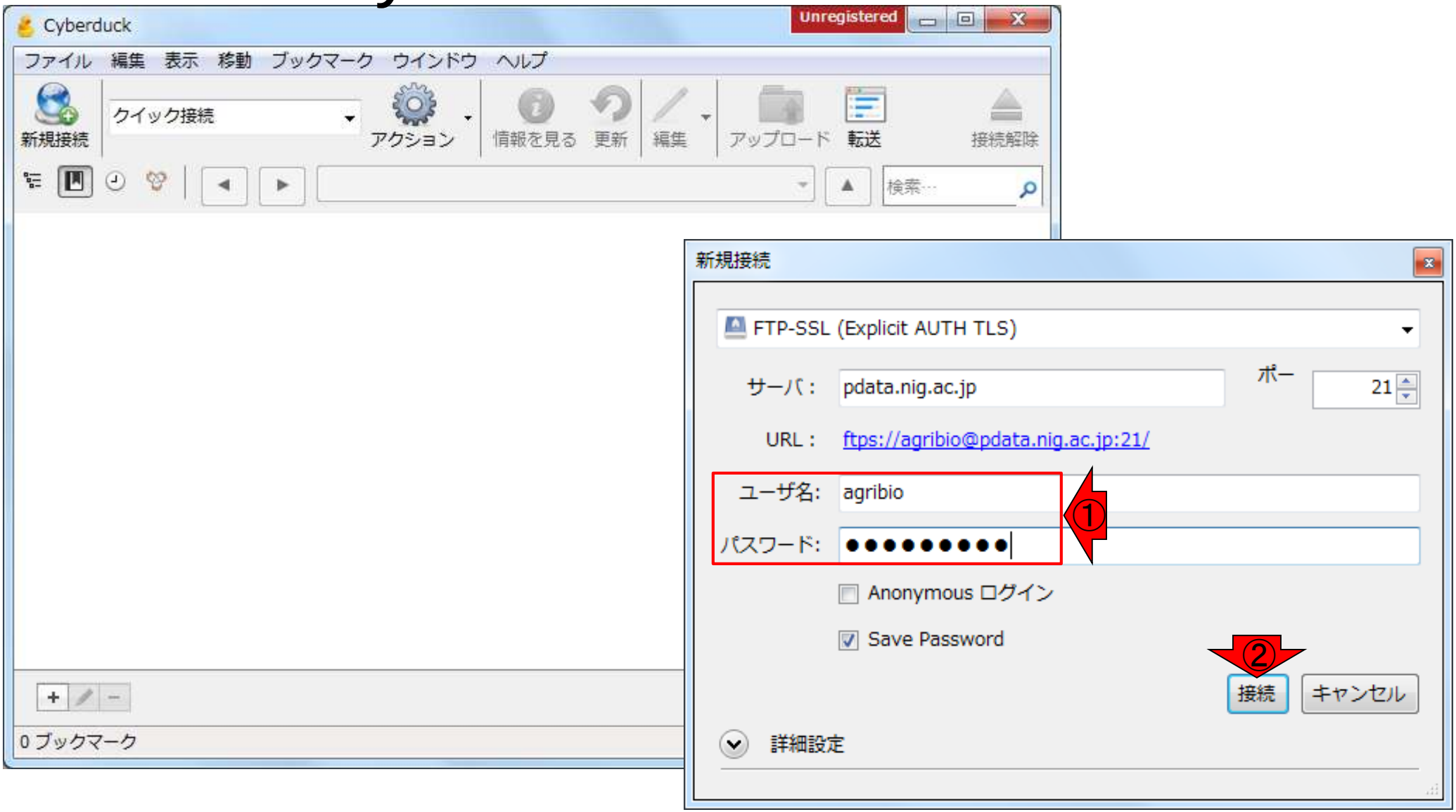
W13-2：Cyberduck設定

第6回W14-2の設定を参考にして①サーバのところにpdata.nig.ac.jpと打つ



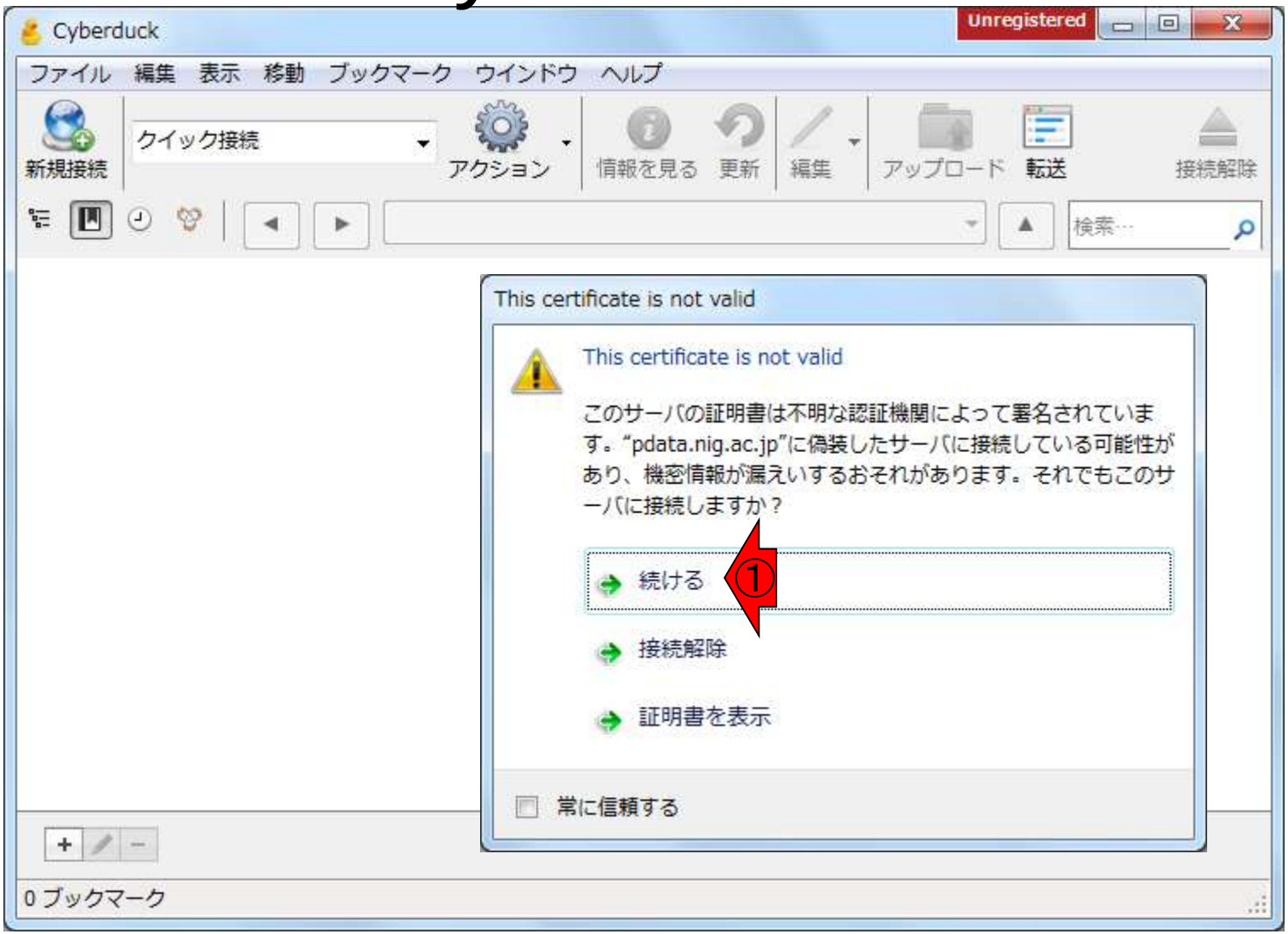
W13-2：Cyberduck設定

①ユーザ名とパスワードはDDBJ Pipelineにログインするときに使うものを入力。②接続



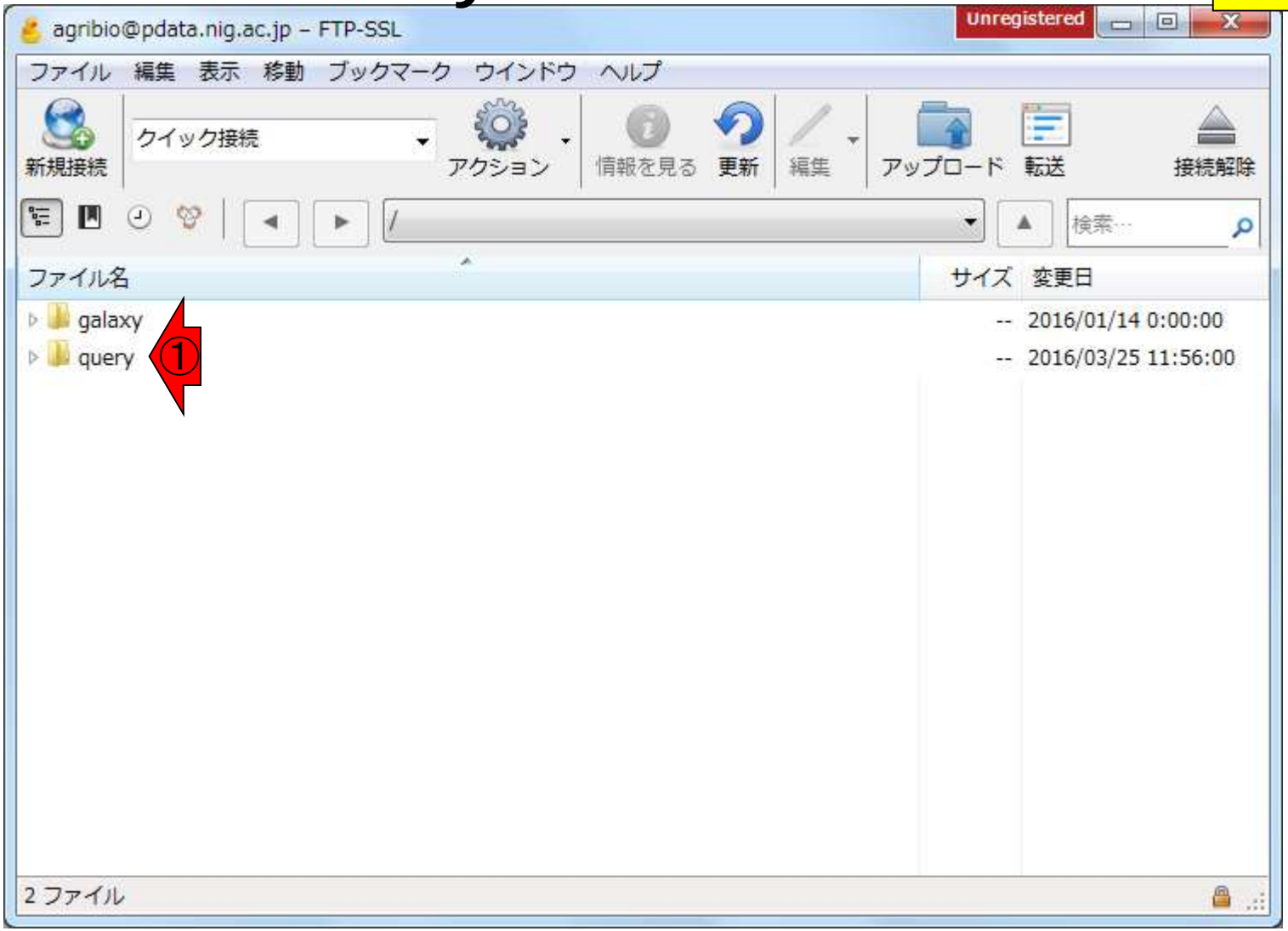
W13-2：Cyberduck設定

①ユーザ名とパスワードはDDBJ Pipelineにログインするときに使うものを入力



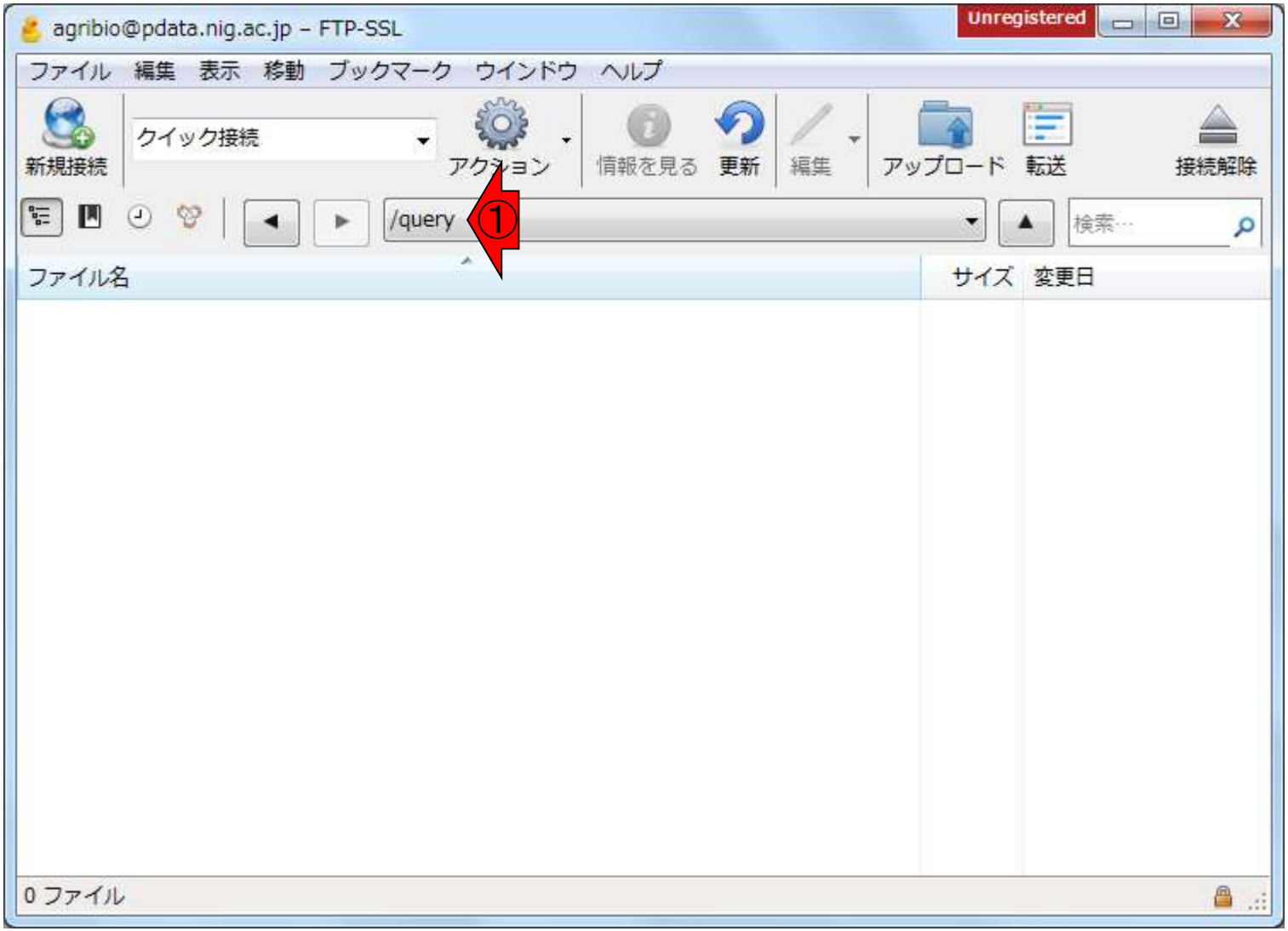
W13-3：Cyberduck接続

第6回W14-4とほぼ同じ状態。
DDBJ Pipelineにアップロード
したいので、①queryをクリック



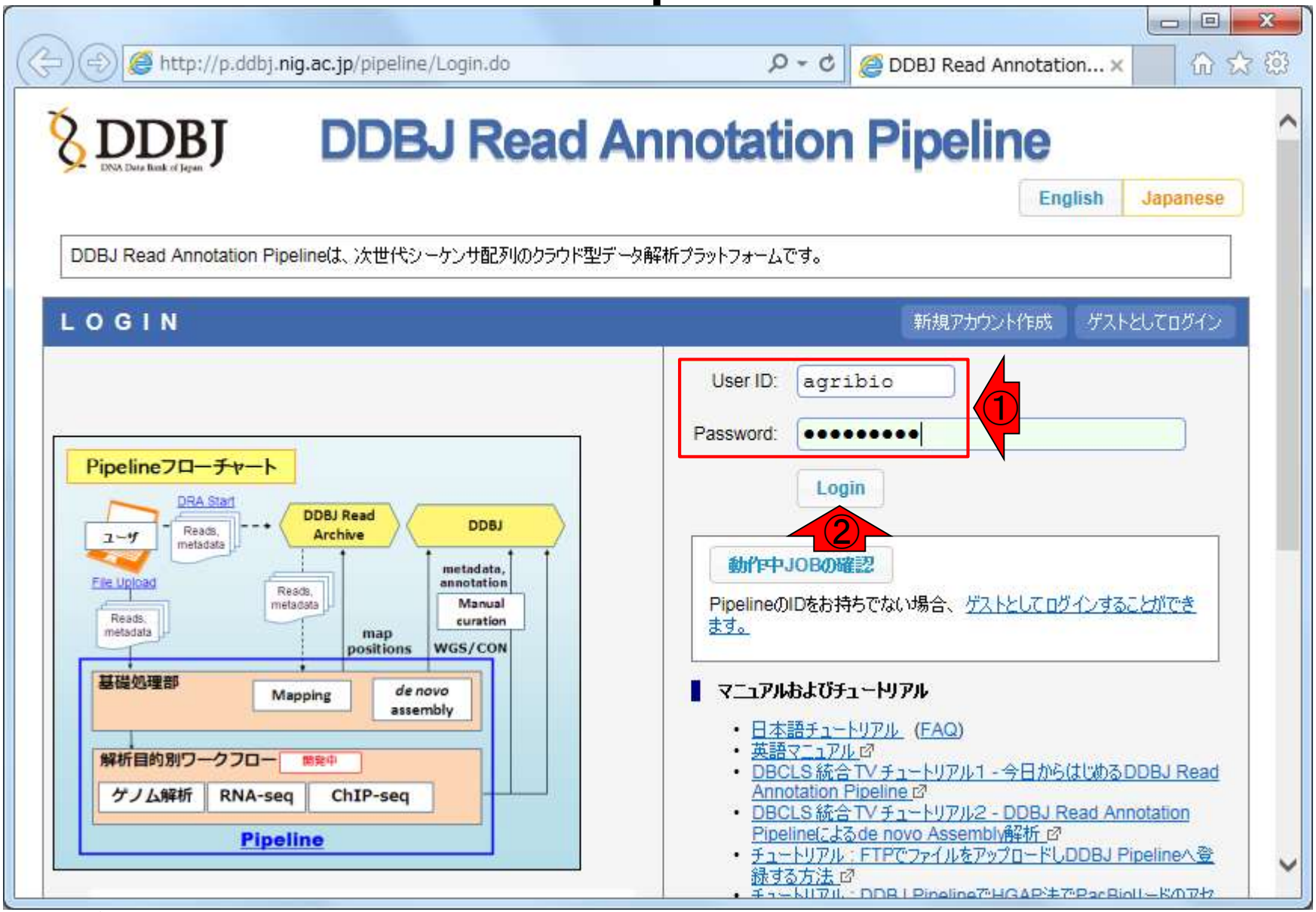
W13-4：アップロード

①queryに移動できていることがわかる。②アップロードしたいファイルをドラッグ&ドロップで移動



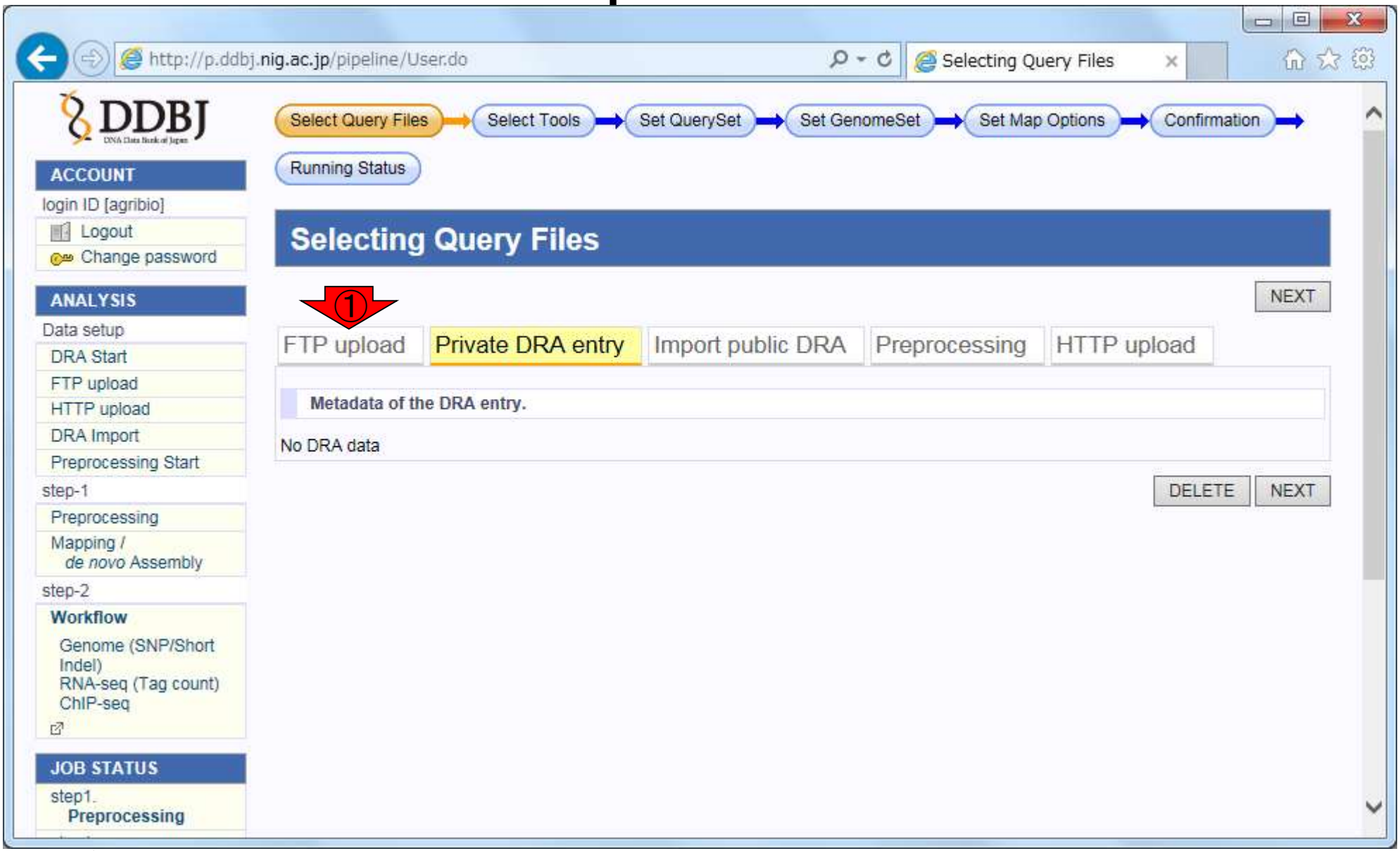
W14-1：DDBJ Pipeline

DDBJ Pipelineの初期画面。①自分のIDとパスワードを入力して、②Login



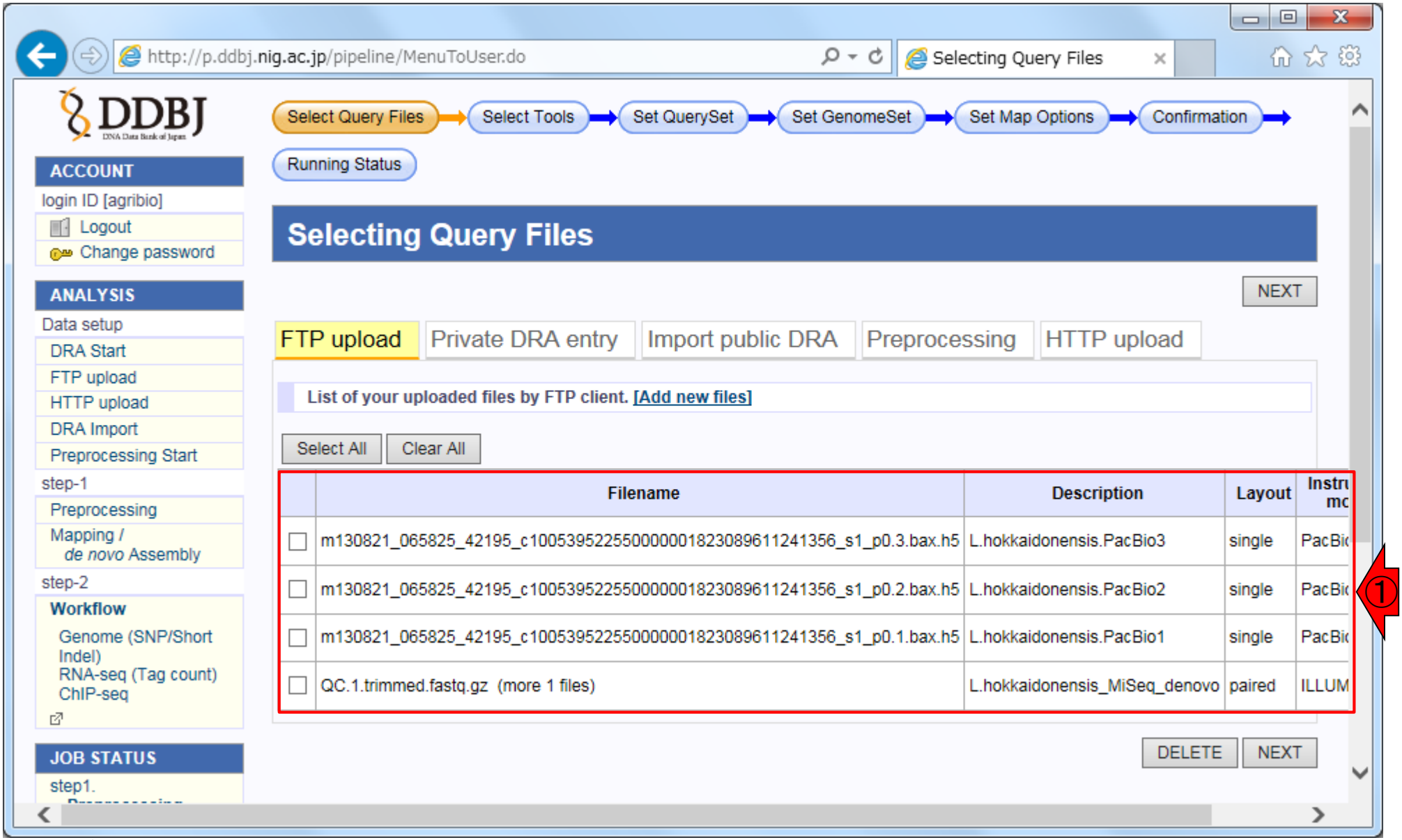
W14-1：DDBJ Pipeline

①FTP upload。マップする側のファイルを指定するのが目的



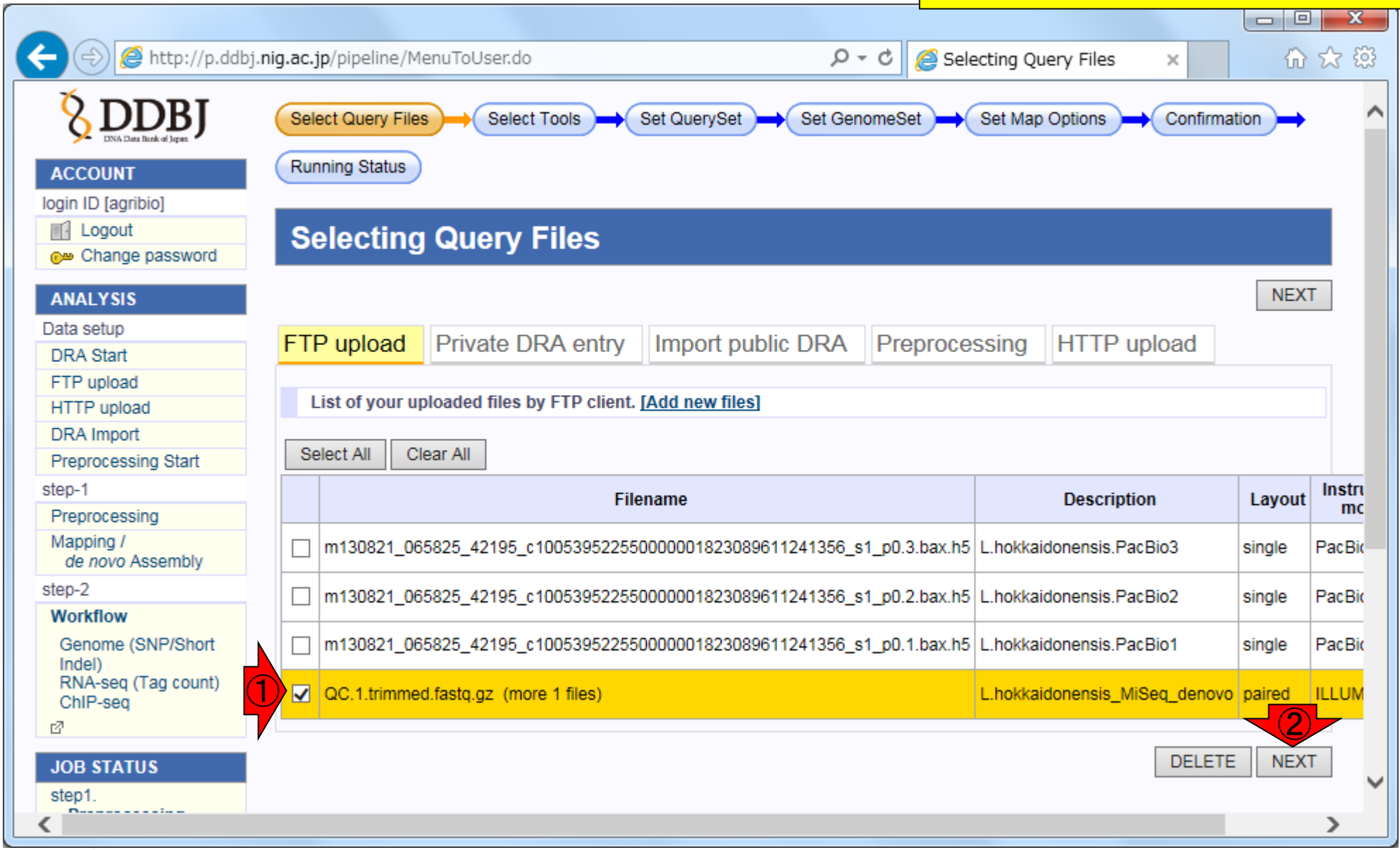
W14-2：マップする側

①赤枠部分の見え方はヒトそれぞれだが、連載第6,7回をやったヒトはこんな感じになっているはず



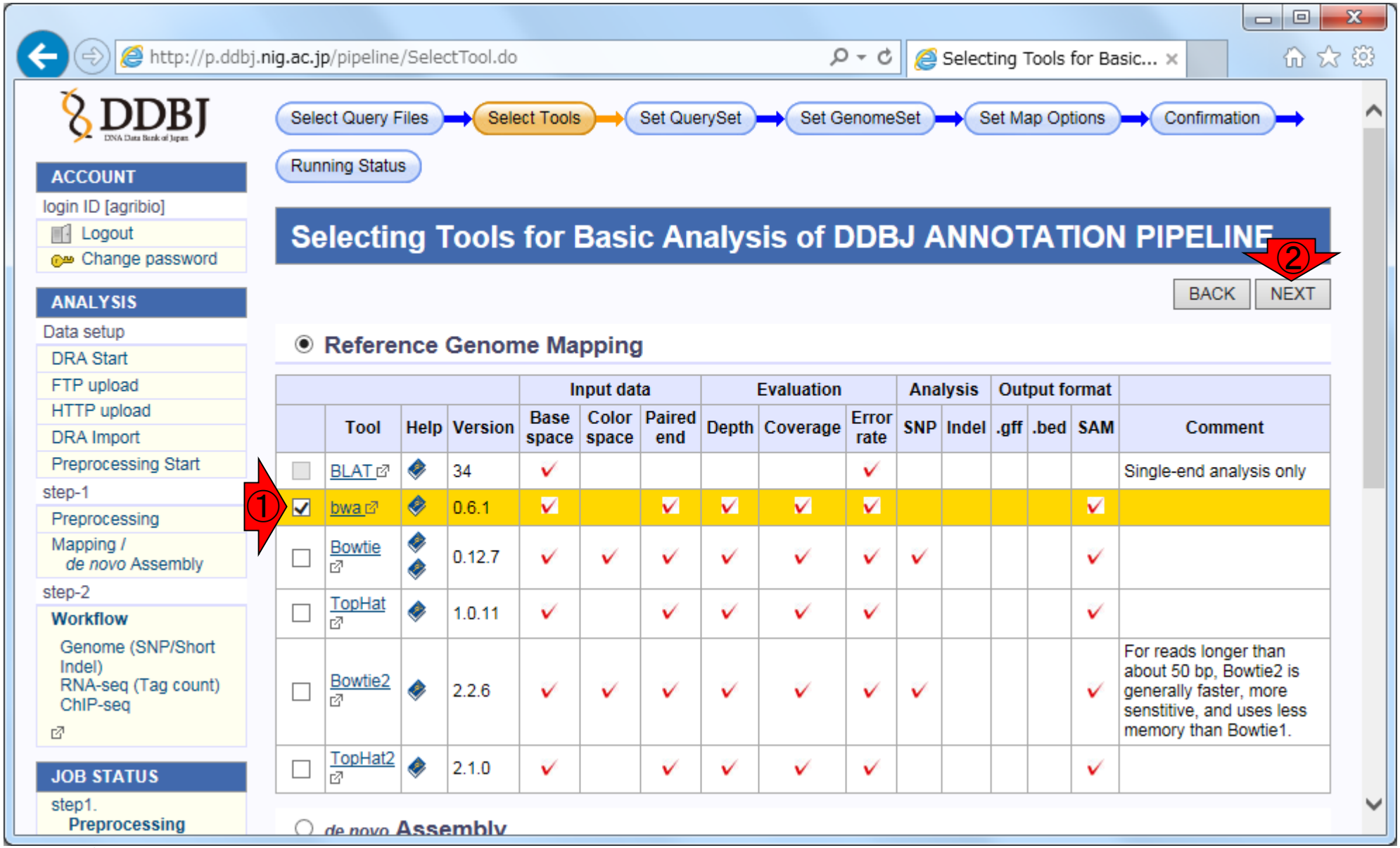
W14-2：マップする側

第6回W14-5あたりまでで登録(registration)した①が残っているヒトはこれにチェックを入れて、②NEXT



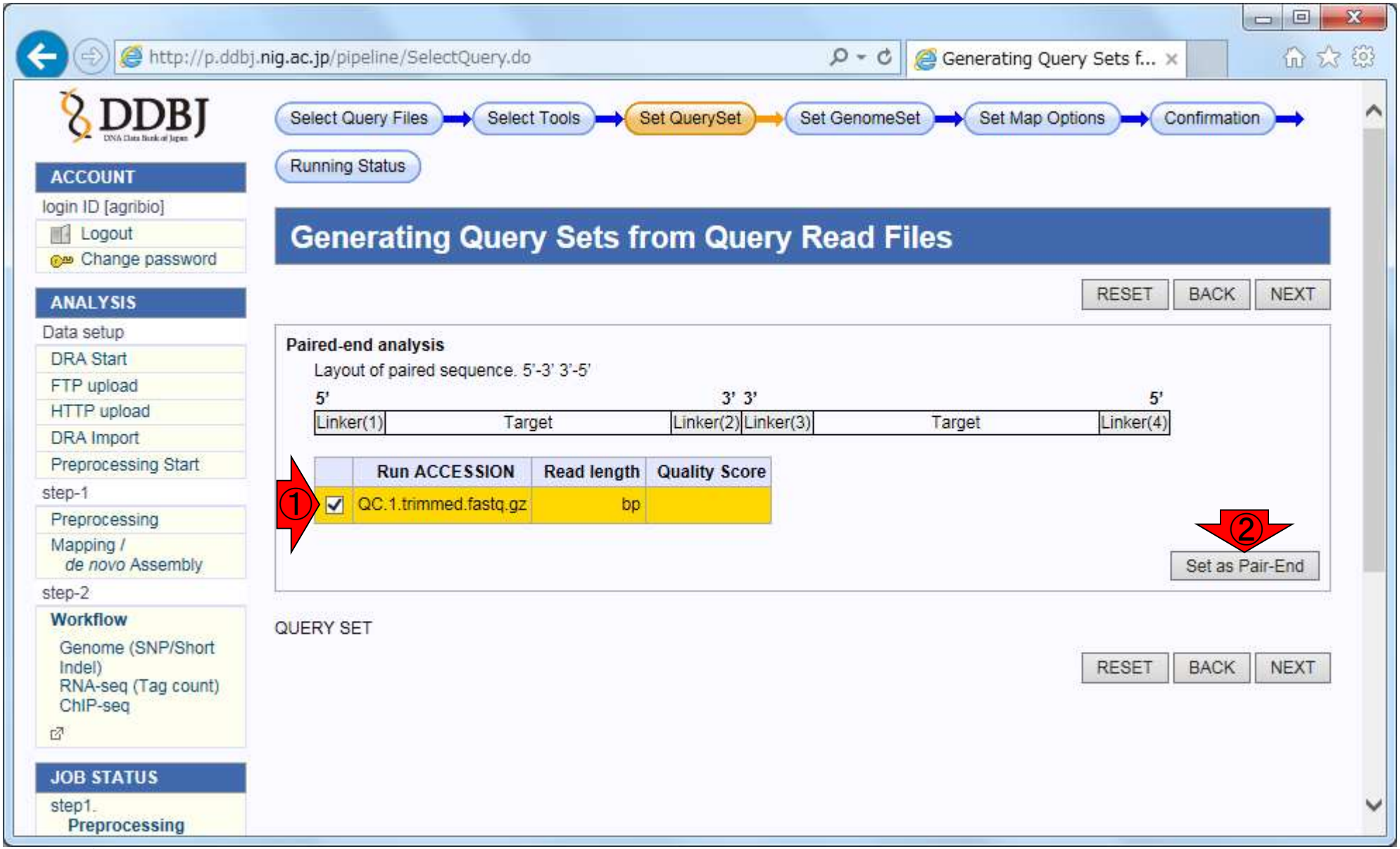
①bwa (ver. 0.6.1)にチェックを入れて、②NEXT

W14-3：bwa



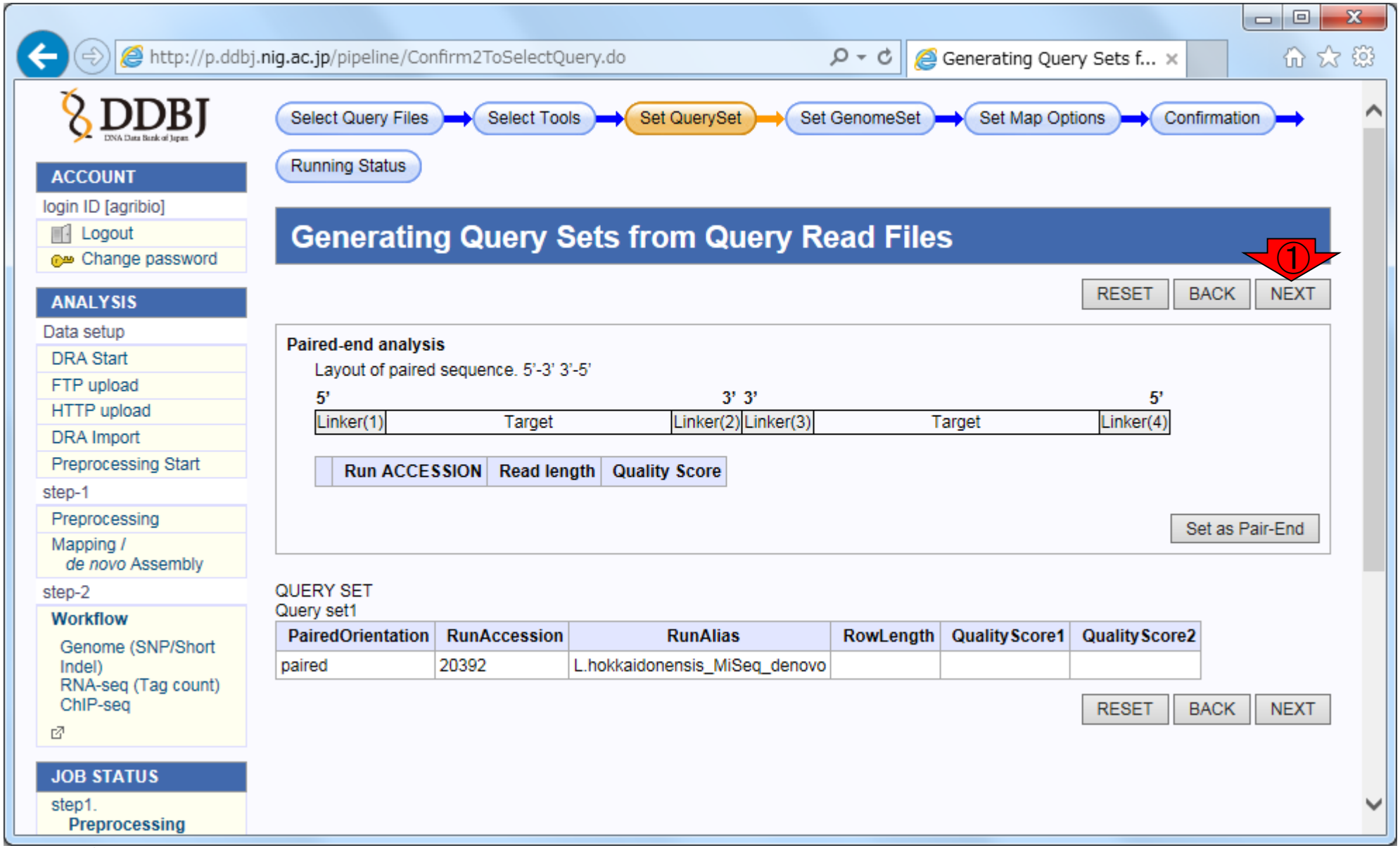
W14-3：bwa

①解析したいデータにチェックをいれて、②Set as Pair-End



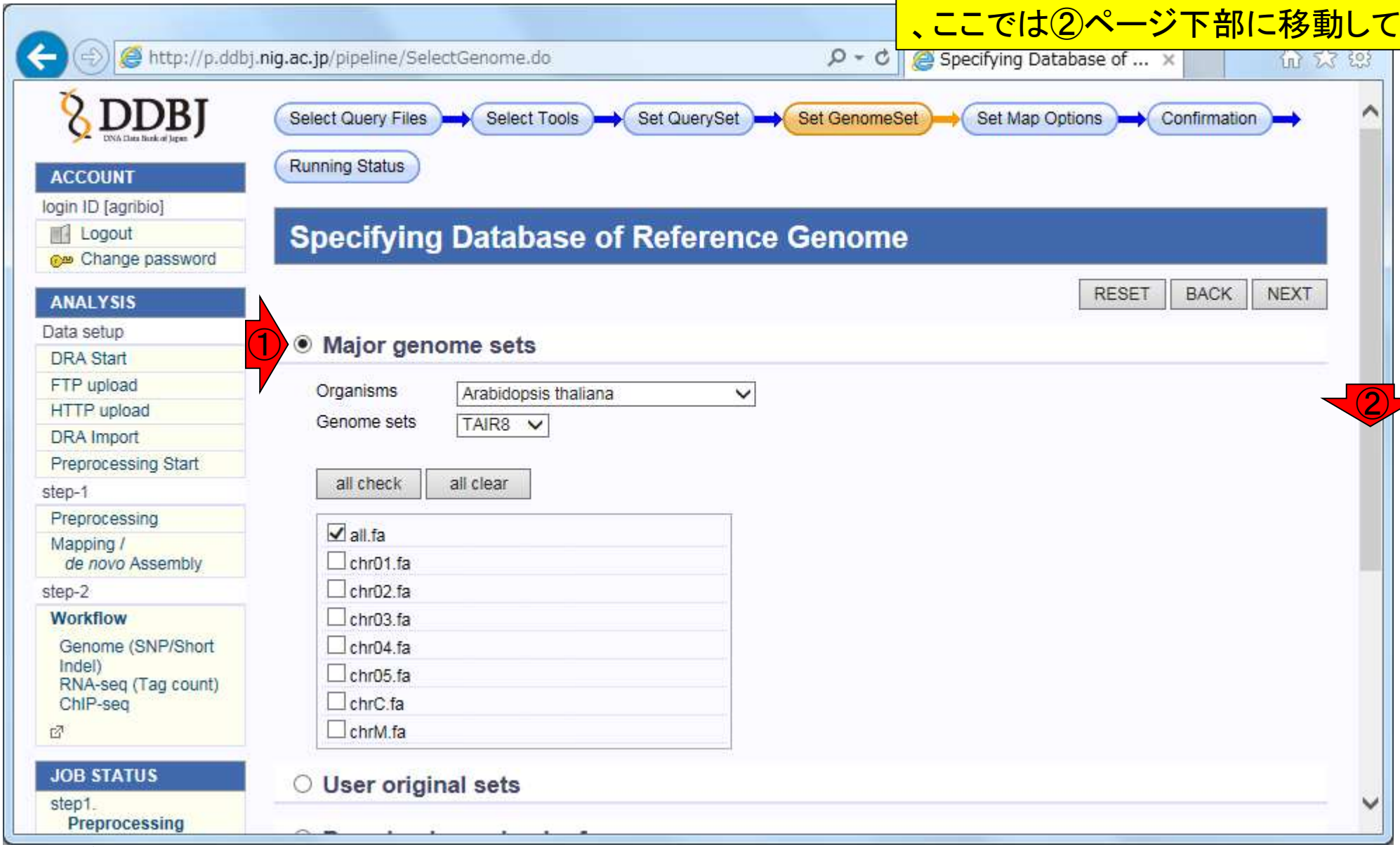
①NEXT

W14-3：bwa



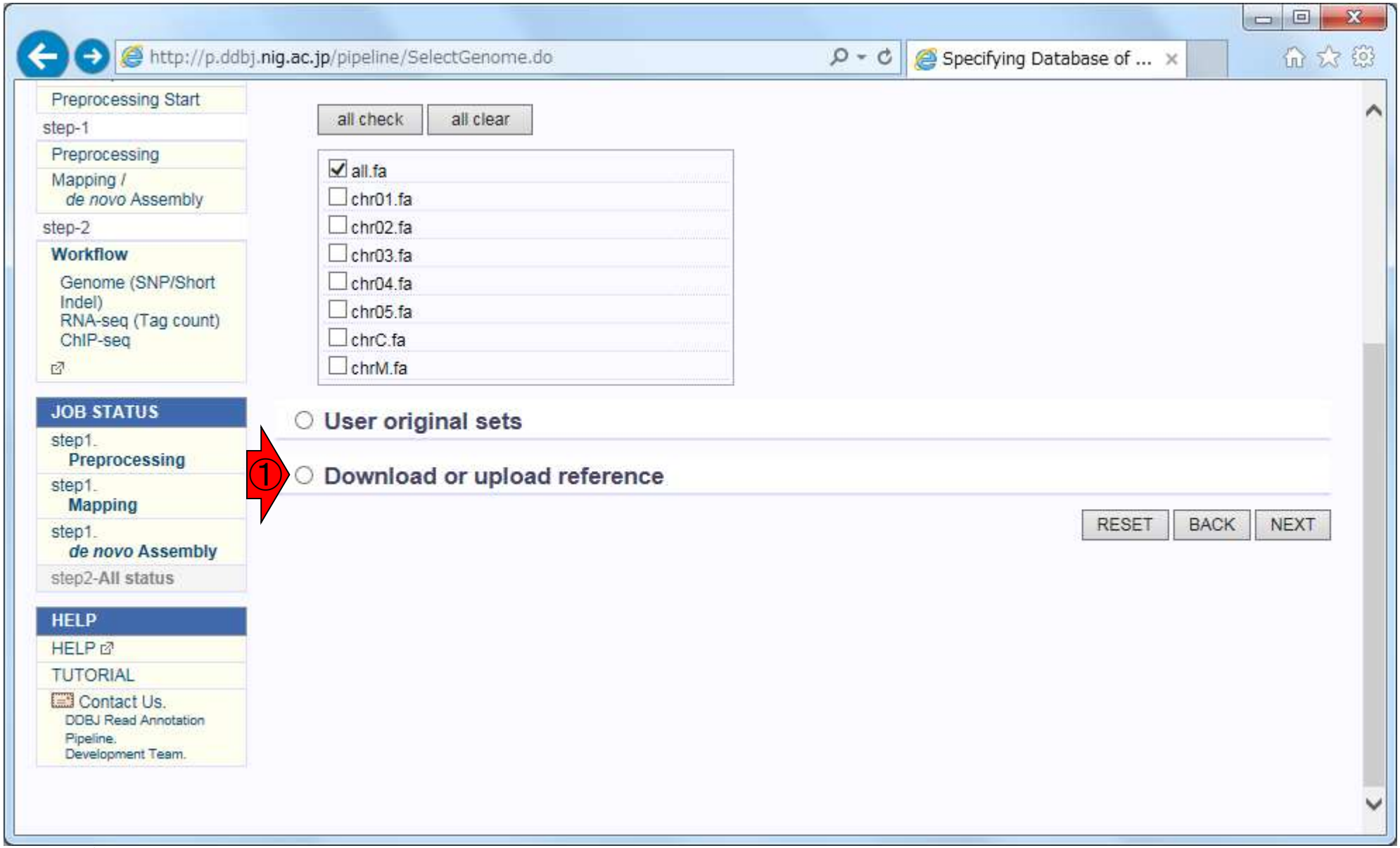
W14-4：マップされる側

マップされる側のリファレンス配列ファイル(LH_draft.fa)を指定。①デフォルトはMajor genome setsになっているが、ここでは②ページ下部に移動して…



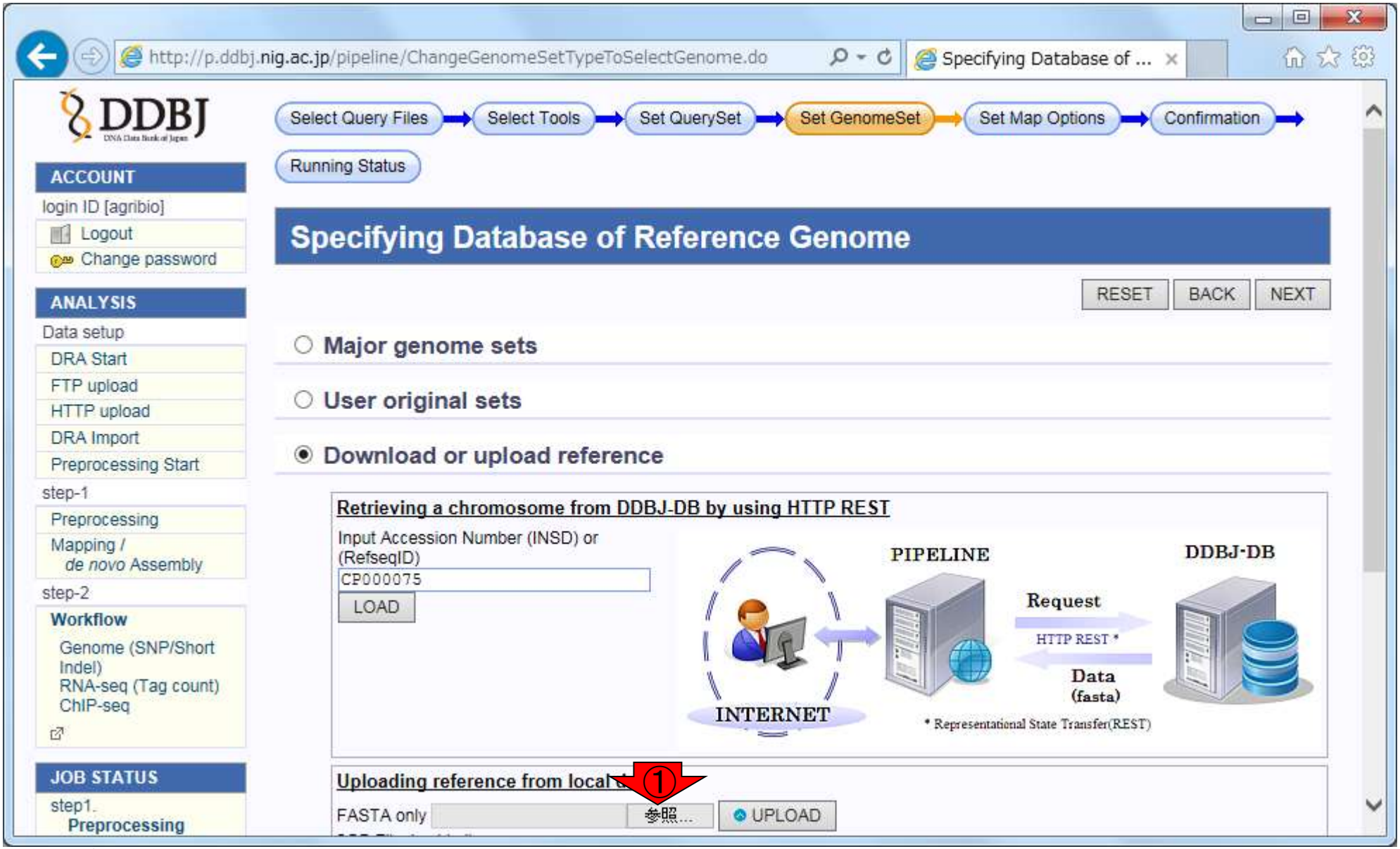
①Download or upload reference

W14-4：マップされる側



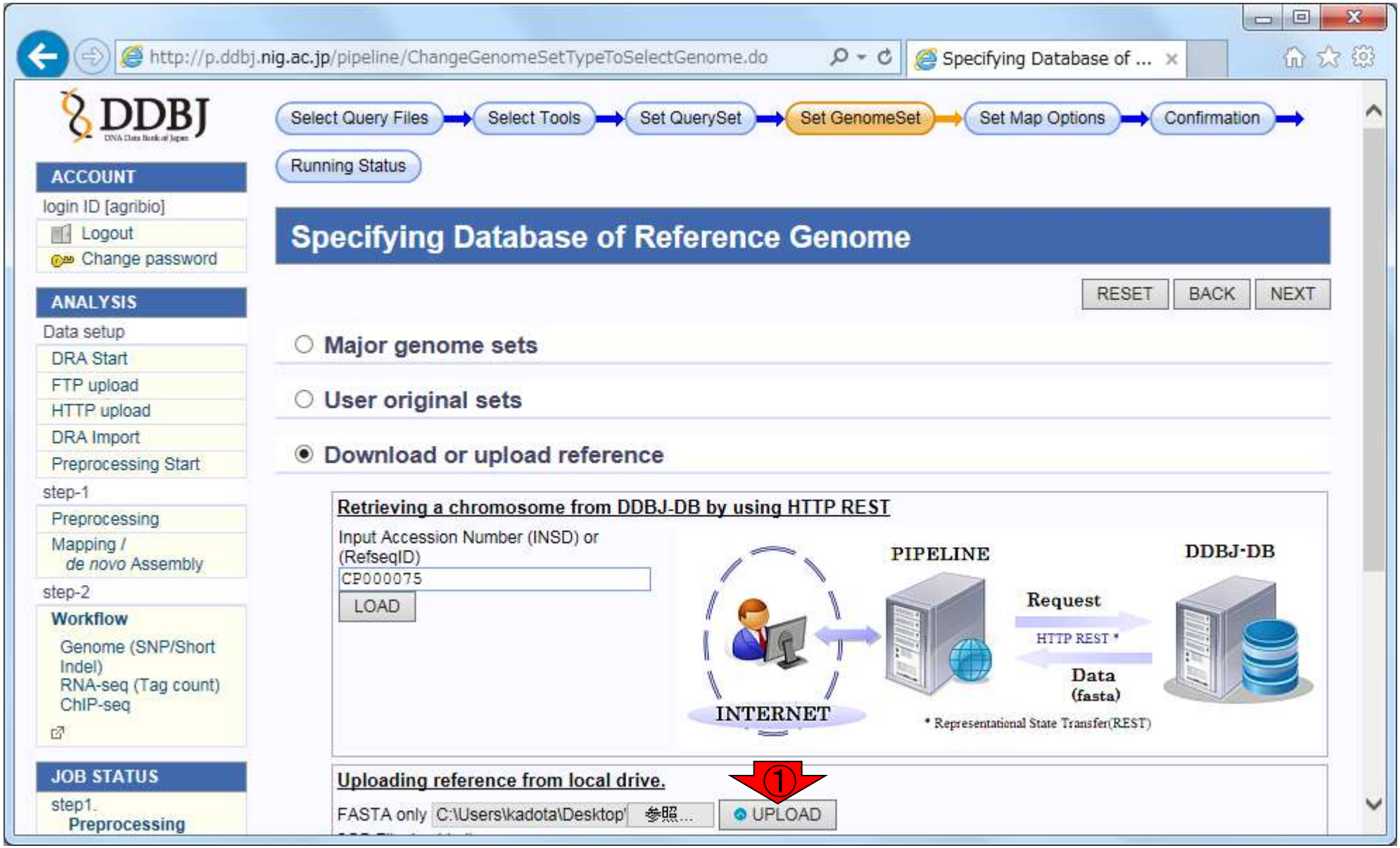
W14-4：マップされる側

①でマップされる側のリファレンス配列ファイル(LH_draft.fa)を指定



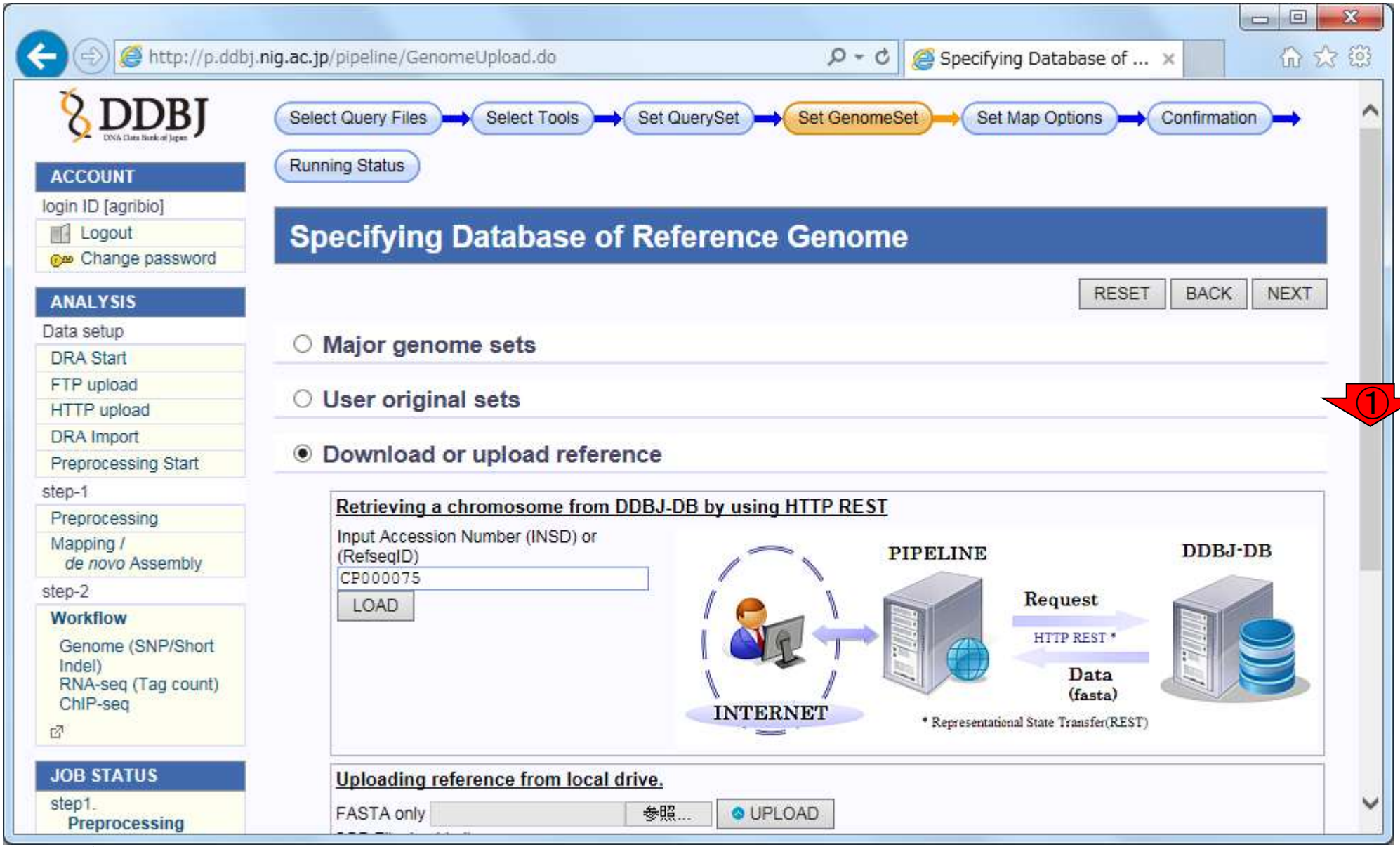
W14-4：マップされる側

マップされる側のリファレンス配列ファイル(LH_draft.fa)を指定し終わったら、①UPLOAD



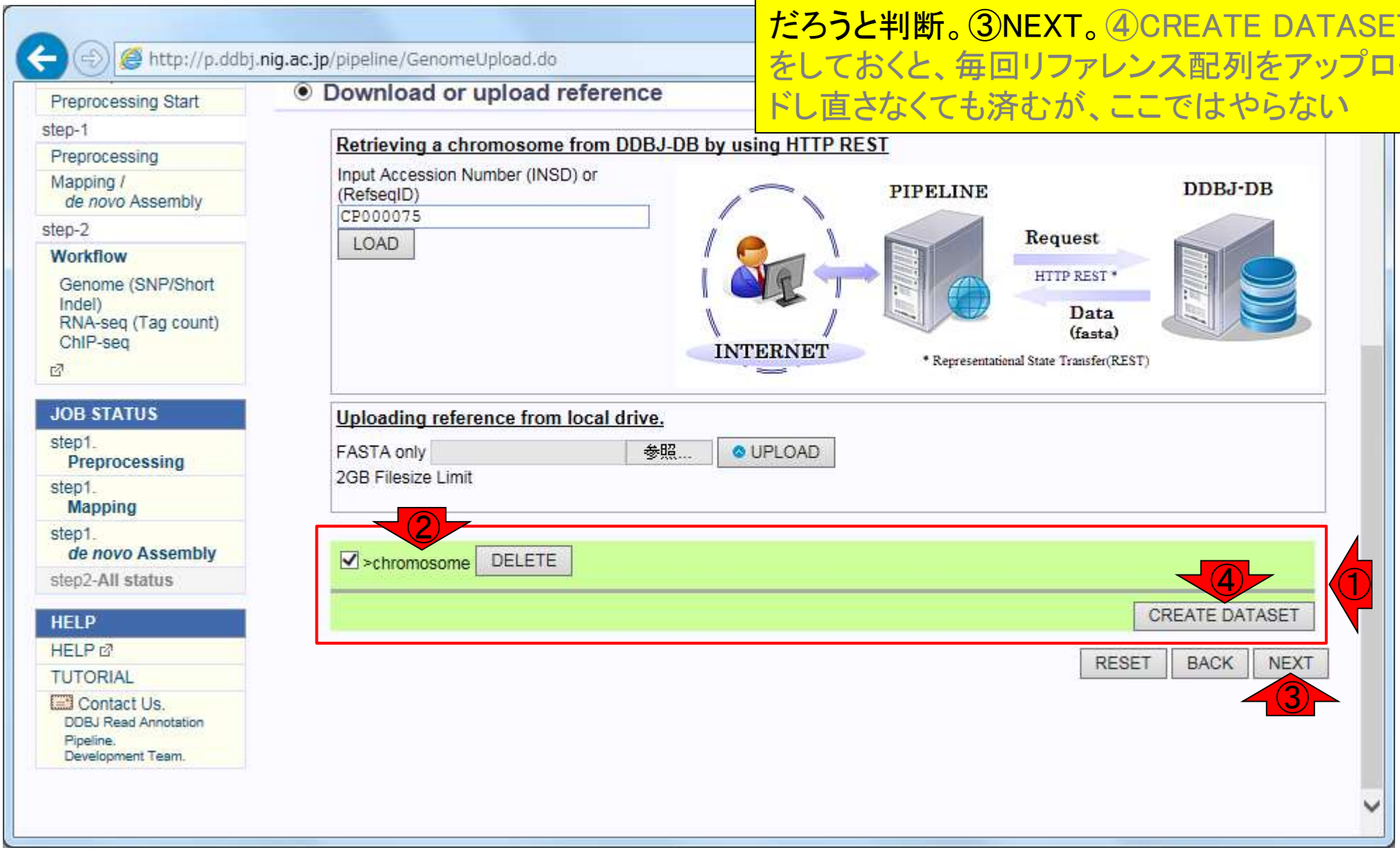
W14-4：マップされる側

特に変化がないが、①ページ下部に移動すると…



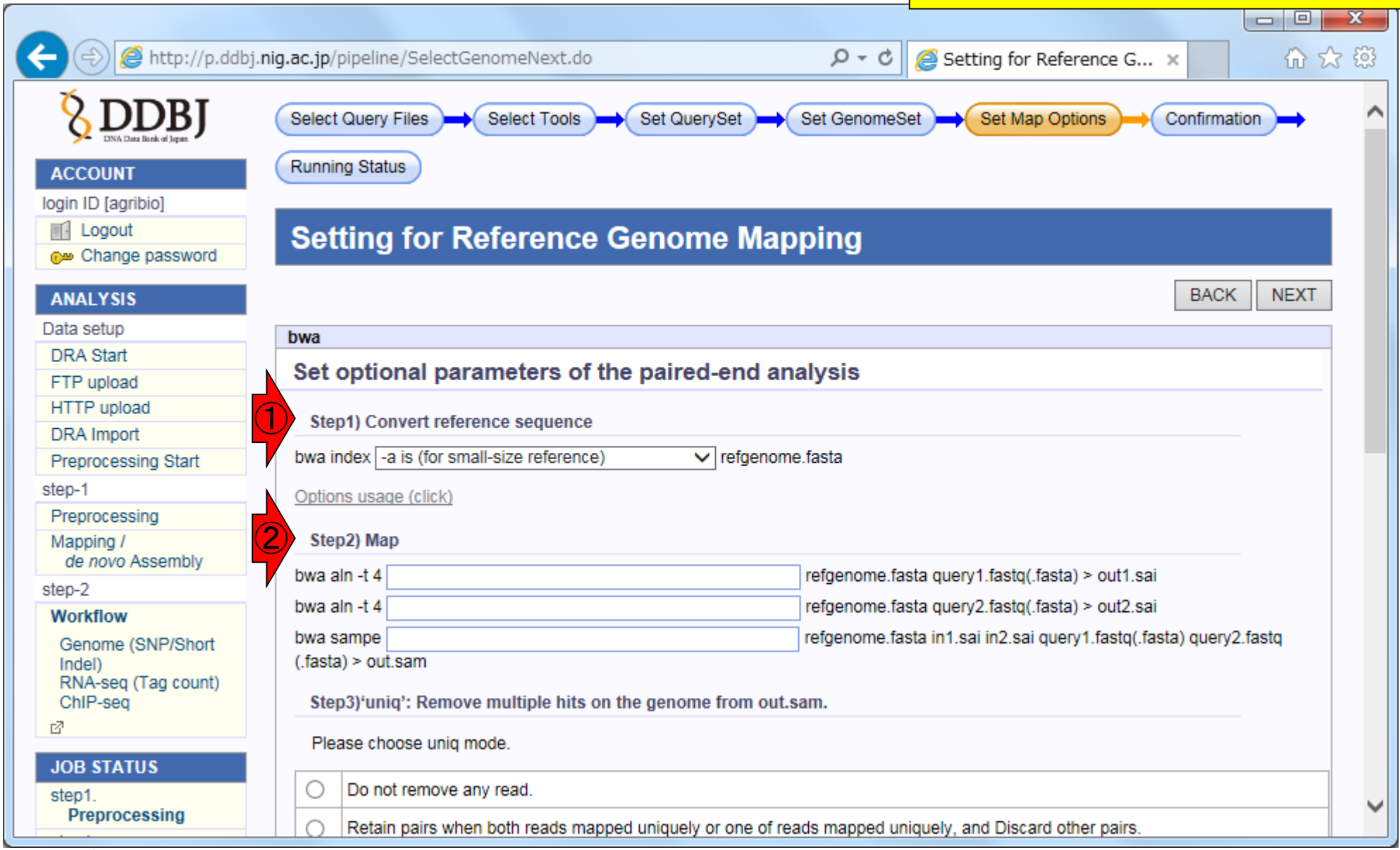
W14-4：マップされる個

①こんなのが見える。②「>chromosome」という文字は、アップロードしたファイルの最初の1行目と同じものなので、うまくアップロードできたのだろうと判断。③NEXT。④CREATE DATASETをしておくと、毎回リファレンス配列をアップロードし直さなくても済むが、ここではやらない



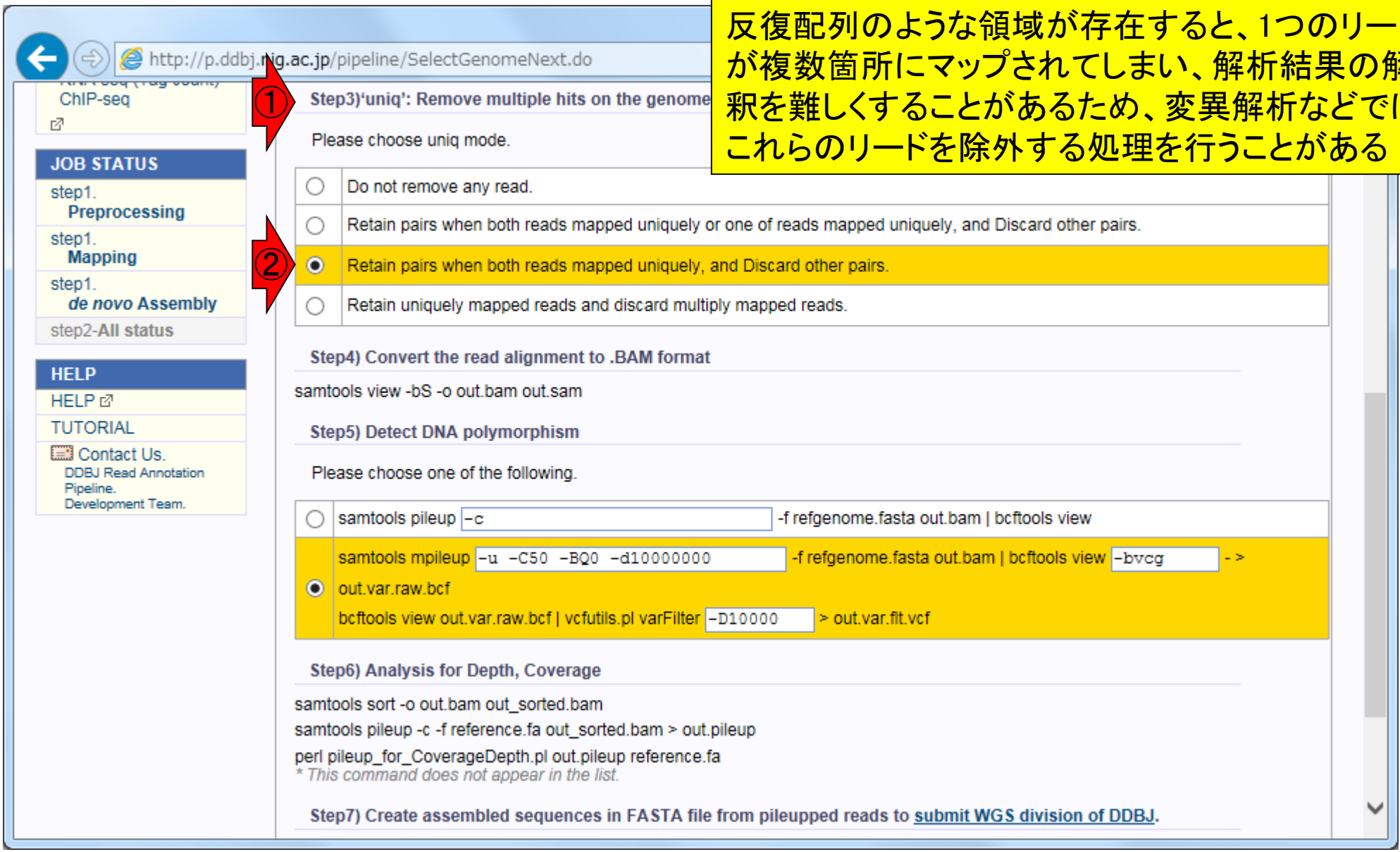
W14-5：オプション指定

基本はデフォルトのままでよい。①Step1）リファレンス配列のインデックス化、②Step2）BWAマッピング本番



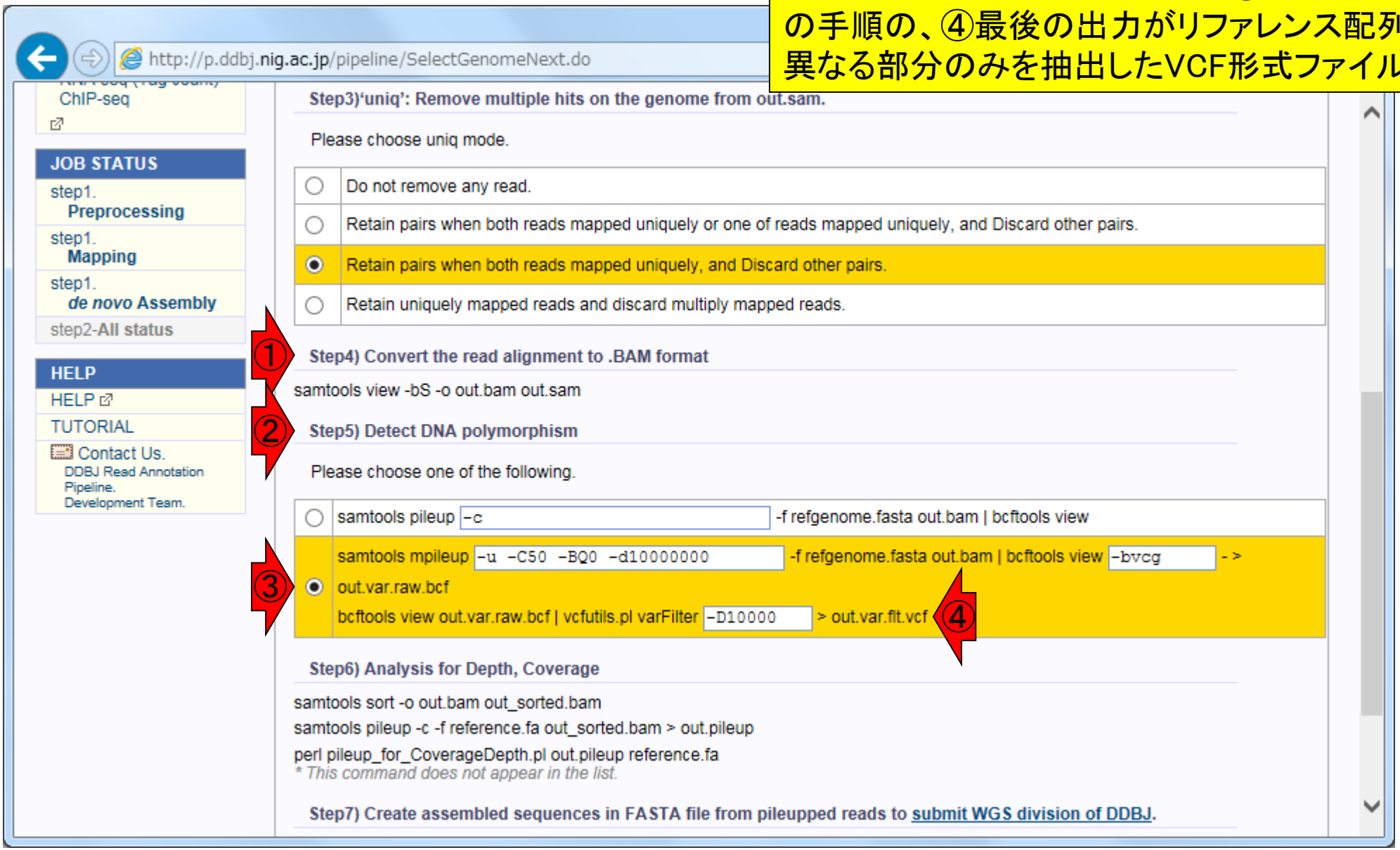
W14-5：オプション指

①Step3) ユニーク化処理。②デフォルトは、ペアのリードの両方がユニークにマップされたものを残し、それ以外のリードは除去する処理を行っている。反復配列のような領域が存在すると、1つのリードが複数箇所にマップされてしまい、解析結果の解釈を難しくすることがあるため、変異解析などではこれらのリードを除外する処理を行うことがある



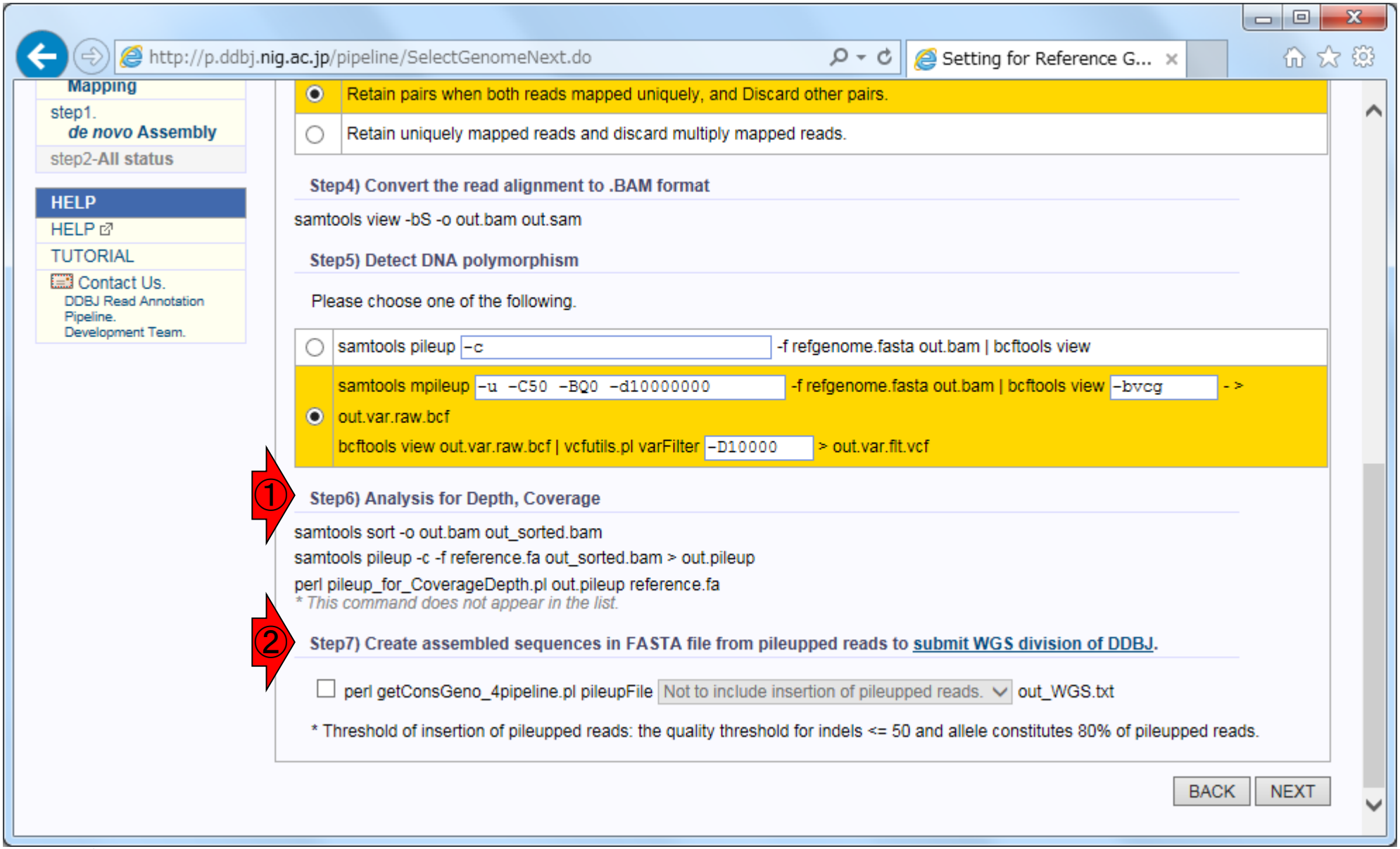
W14-5：オプション指定

①Step4）BWAの出力であるSAMファイルを入力として、BAMファイルを作成。②Step5）DNA多型の検出（変異解析のこと）。③デフォルトの手順の、④最後の出力がリファレンス配列と異なる部分のみを抽出したVCF形式ファイル



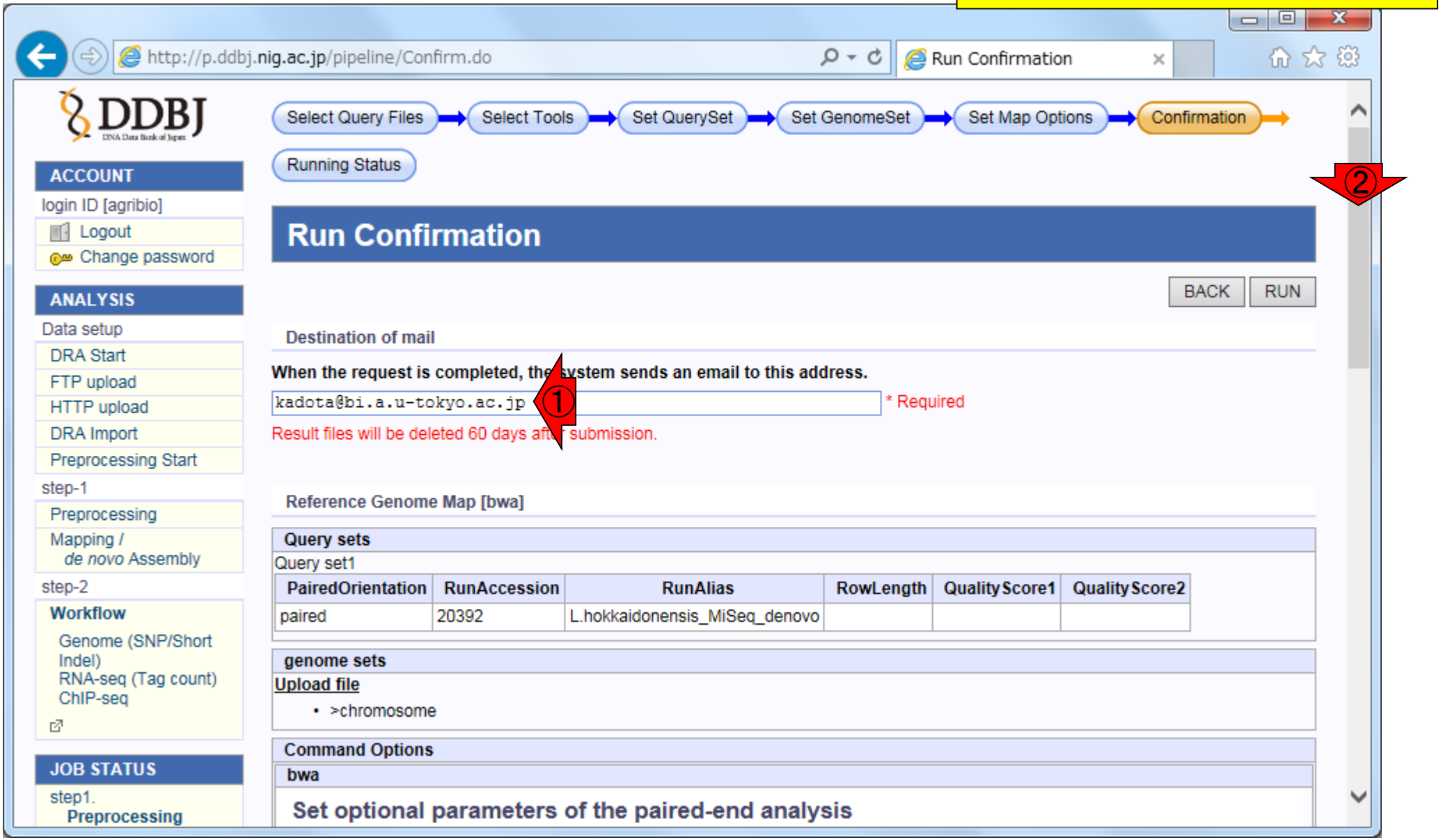
W14-5：オプション指定

①Step6)や、②Step7)は、第8回内容とは関係が薄いので省略



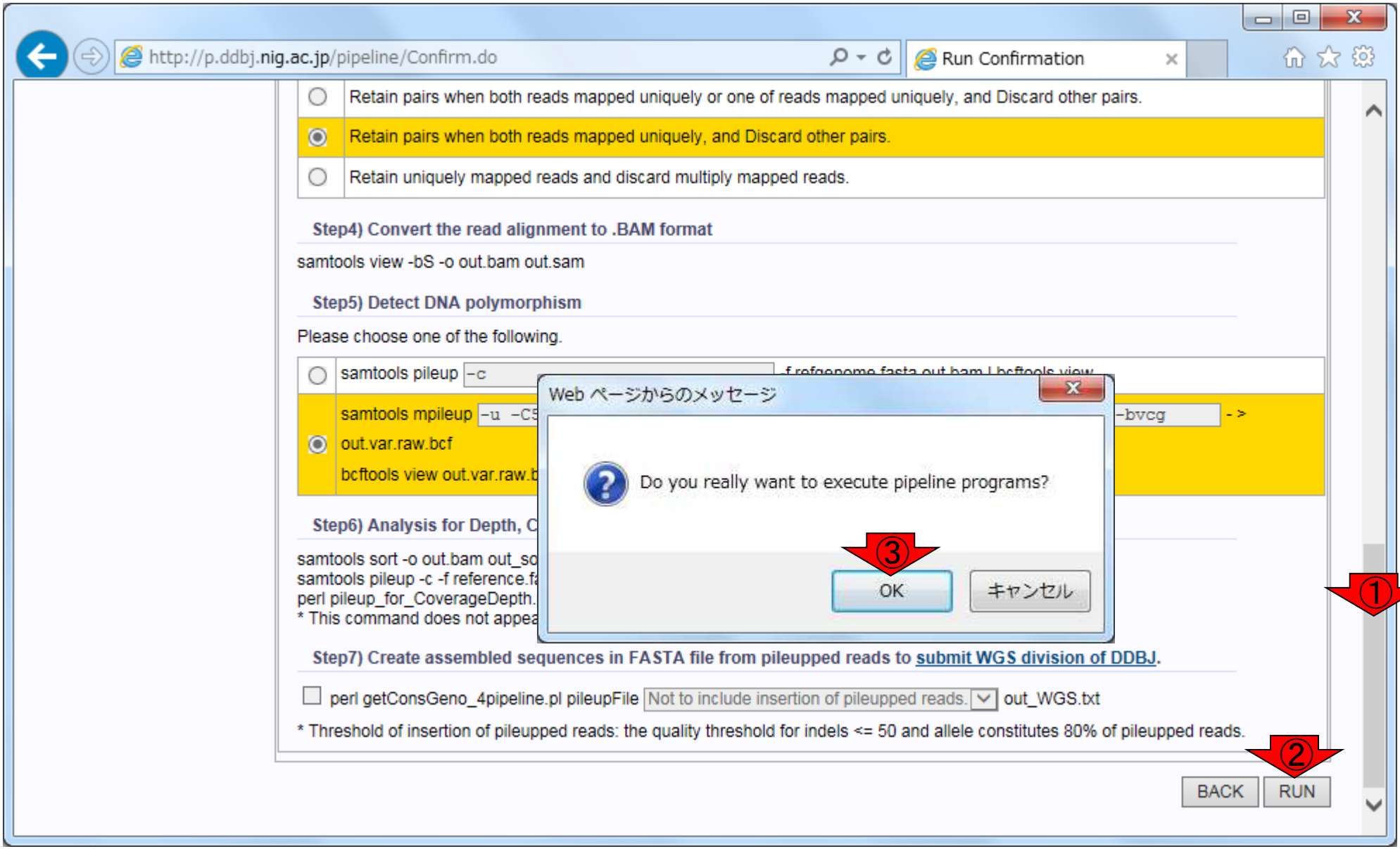
W14-6：確認して実行

①ジョブ完了後のメール送信先を指定して、②一応ページ下部までざっと眺める



W14-6：確認して実行

①一応ページ下部までざっと眺めて問題ないようなら、②RUN、③OK



W14-6：確認して実行

ジョブ投入完了。メール通知があるまで思考停止

メール通知がきたので結果を見にいく

W15-1：計算終了

2016/09/12 (月) 17:38

pipeline_team@g.nig.ac.jp

Job finished : DDBJ Read Annotation Pipeline

宛先 kadota@bi.a.u-tokyo.ac.jp

ＣＣ pipeline_report@g.nig.ac.jp

このメッセージから余分な改行を削除しました。

Dear agribio,

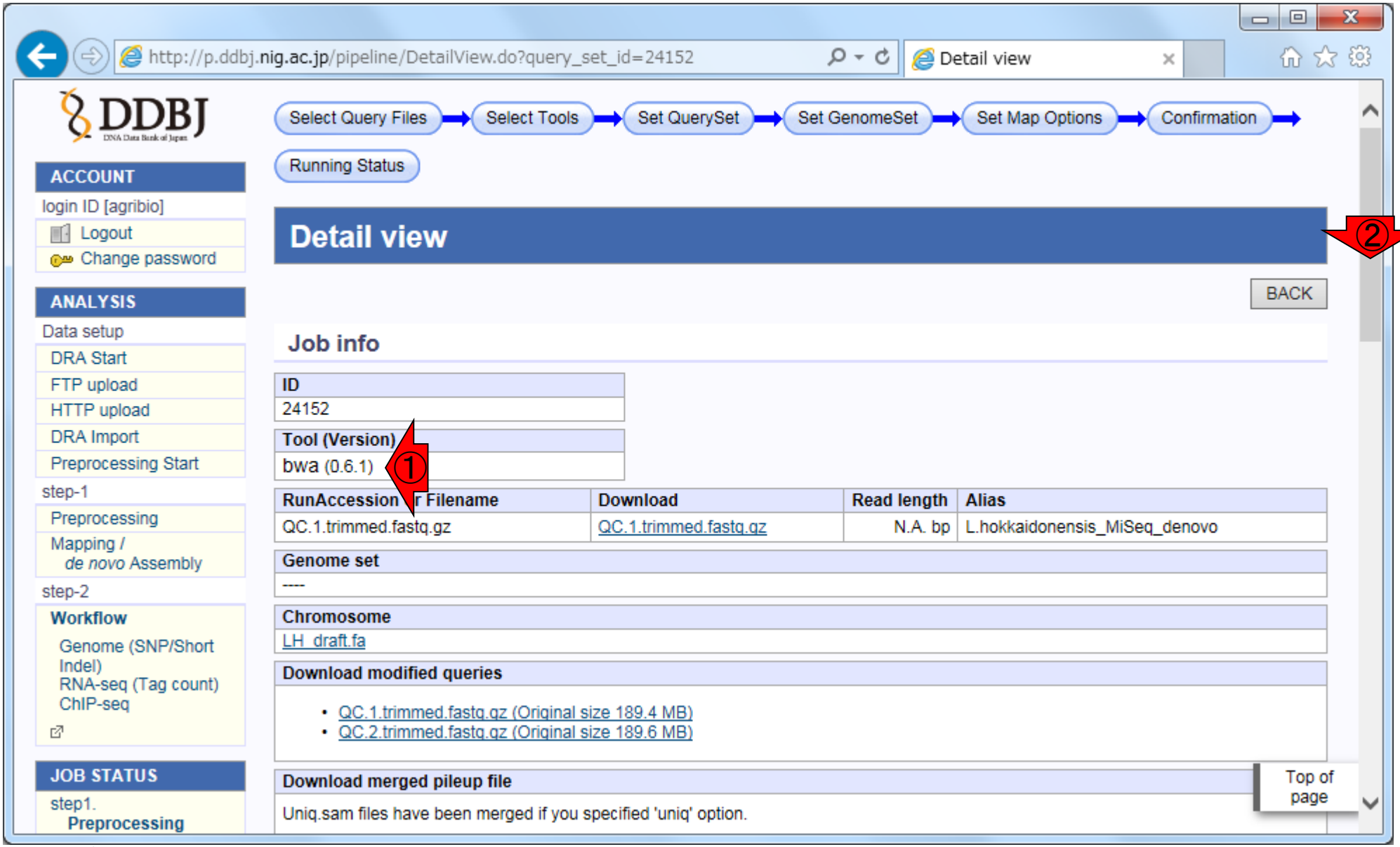
Your request to DDBJ pipeline service has finished.
Please visit the web site to obtain analytical results.
Request ID: 24152
URL: https://p.ddbj.nig.ac.jp/

If you have troubles in this service, please write to pipeline_dev@ddbj.nig.ac.jp Thank you for trying our analytical service.

Regards,
DDBJ

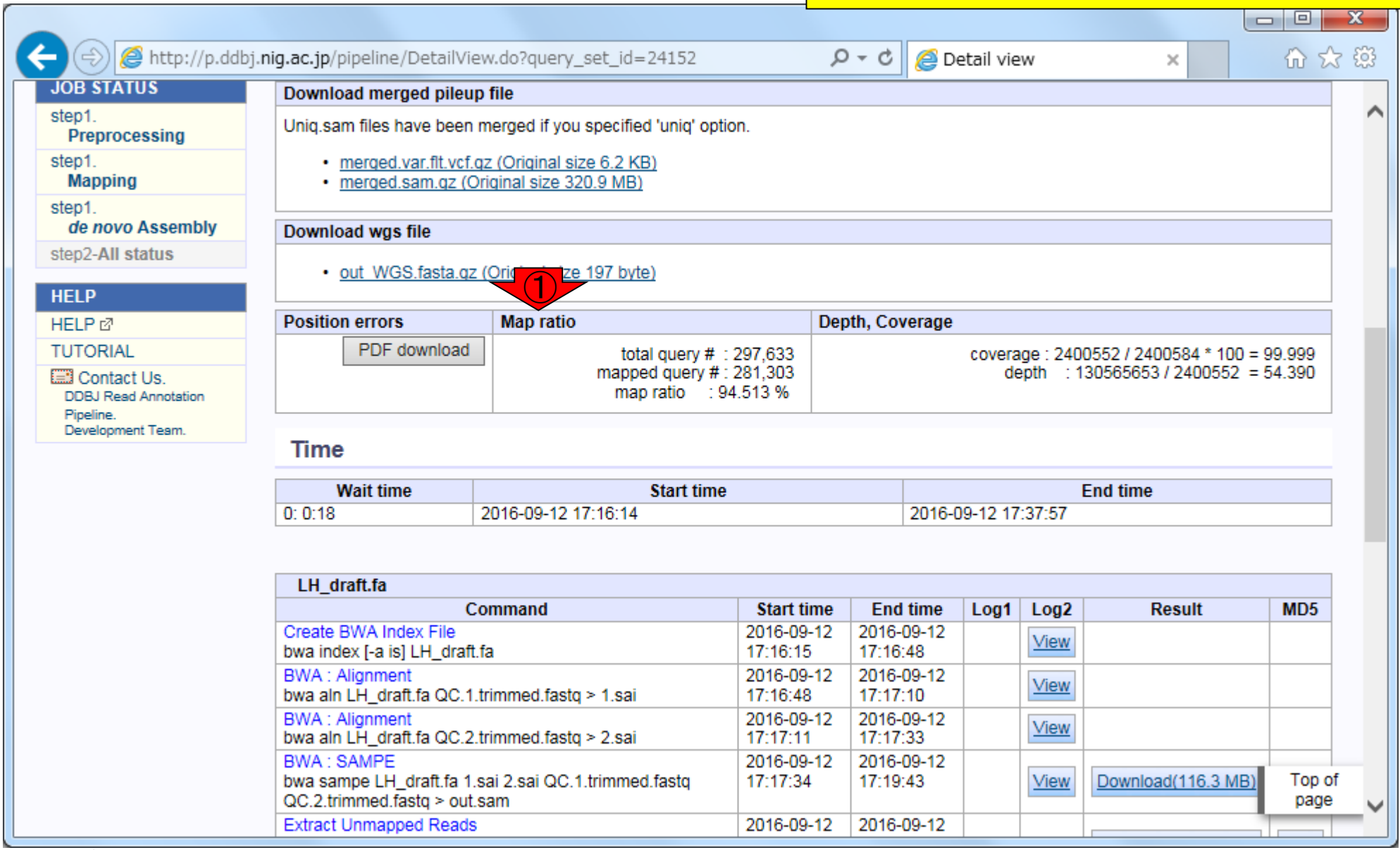
W15-2：結果の要約

①用いたプログラムはbwa（ver. 0.6.1）。②ページ下部に移動



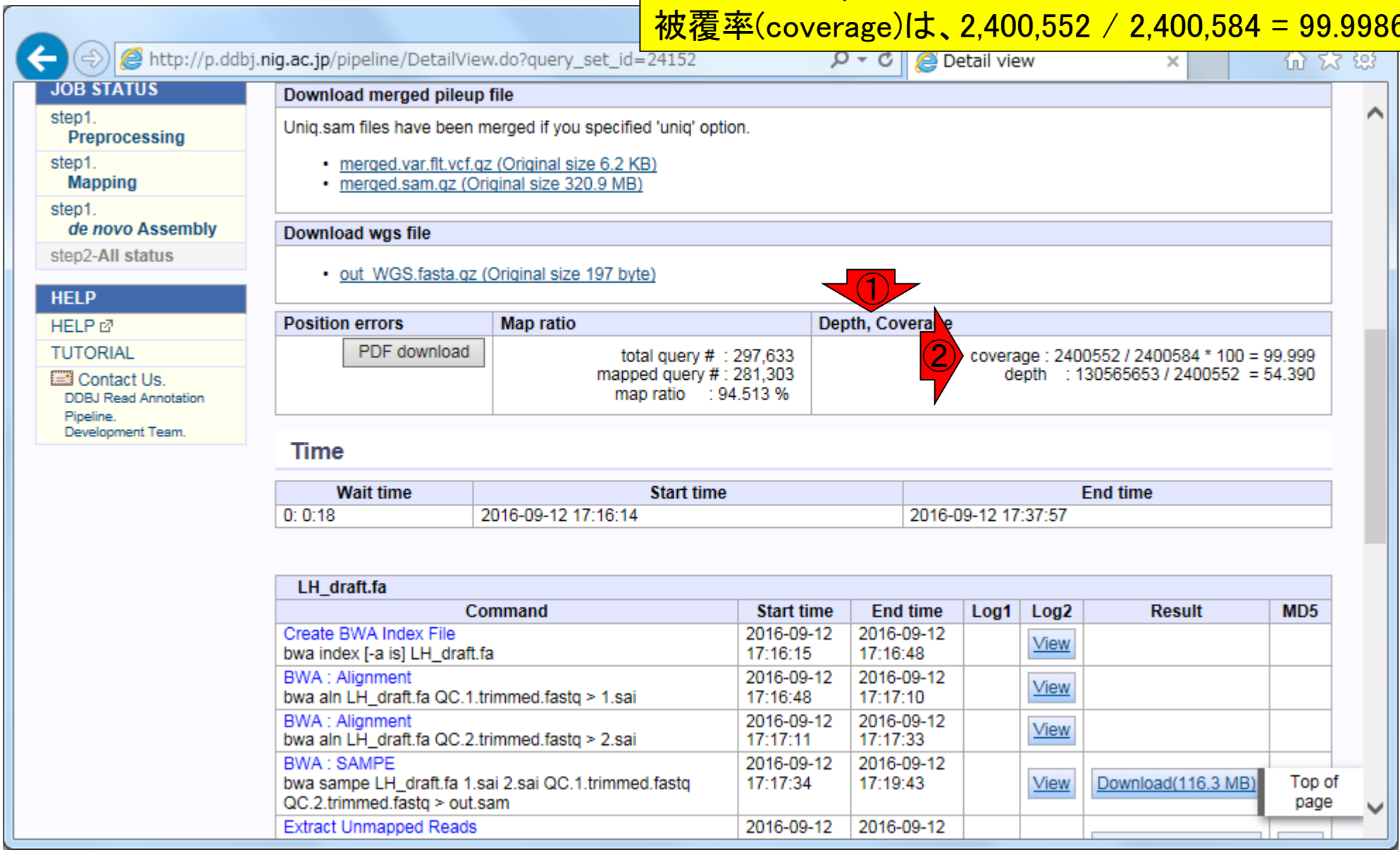
W15-2：結果の要約

①Map ratioのところの解説。全297,633リード中、281,303リードがペアでマップされ、マップ率は281,303/297,633 = 94.513%であった



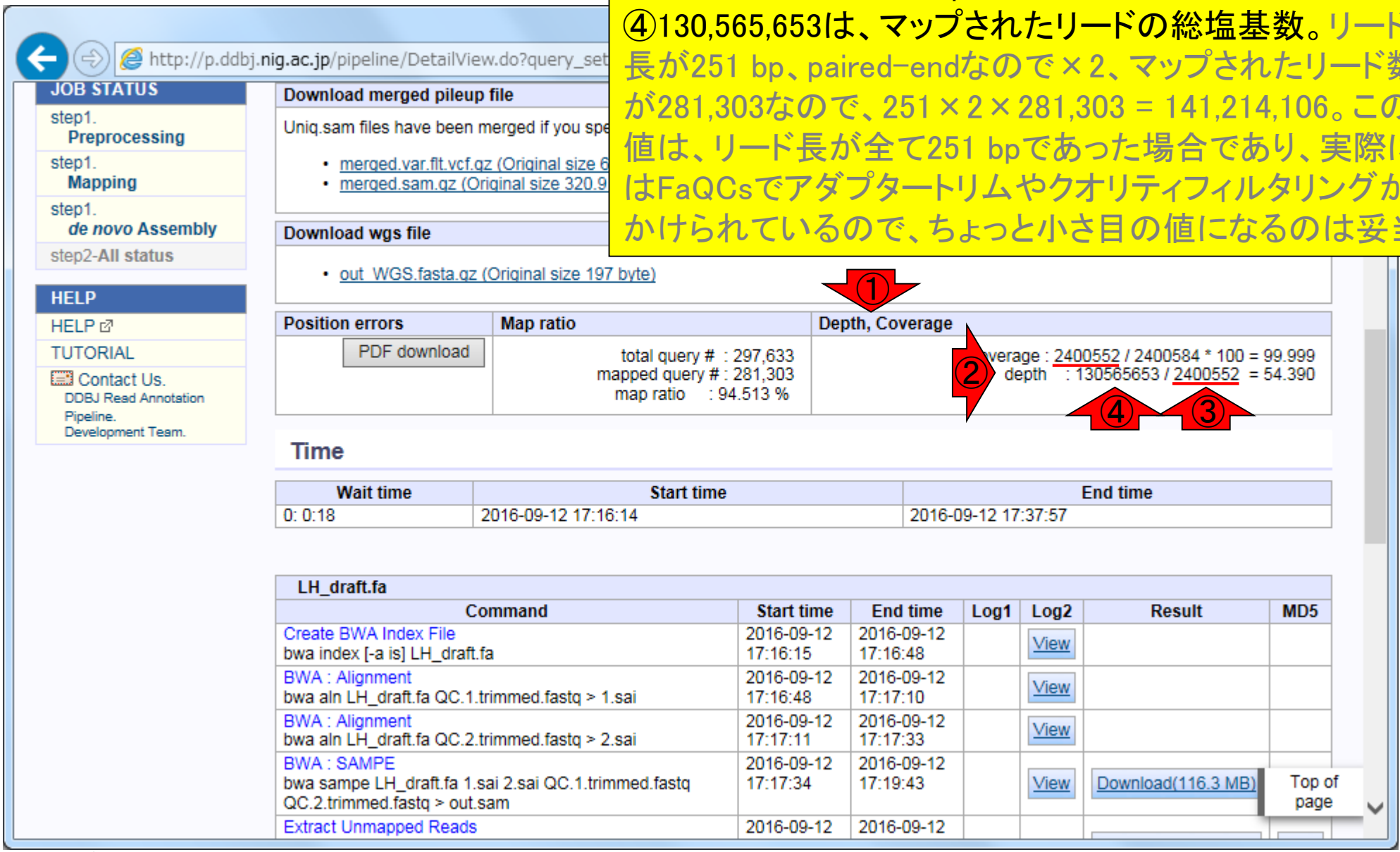
W15-2：結果の要約

①Depth, Coverageのところの解説で、まずは②Coverage。ゲノムサイズ（全長）が2,400,584 bp。そのうち2,400,552 bp分がマップされたリードで覆われている。被覆率(coverage)は、2,400,552 / 2,400,584 = 99.99867%



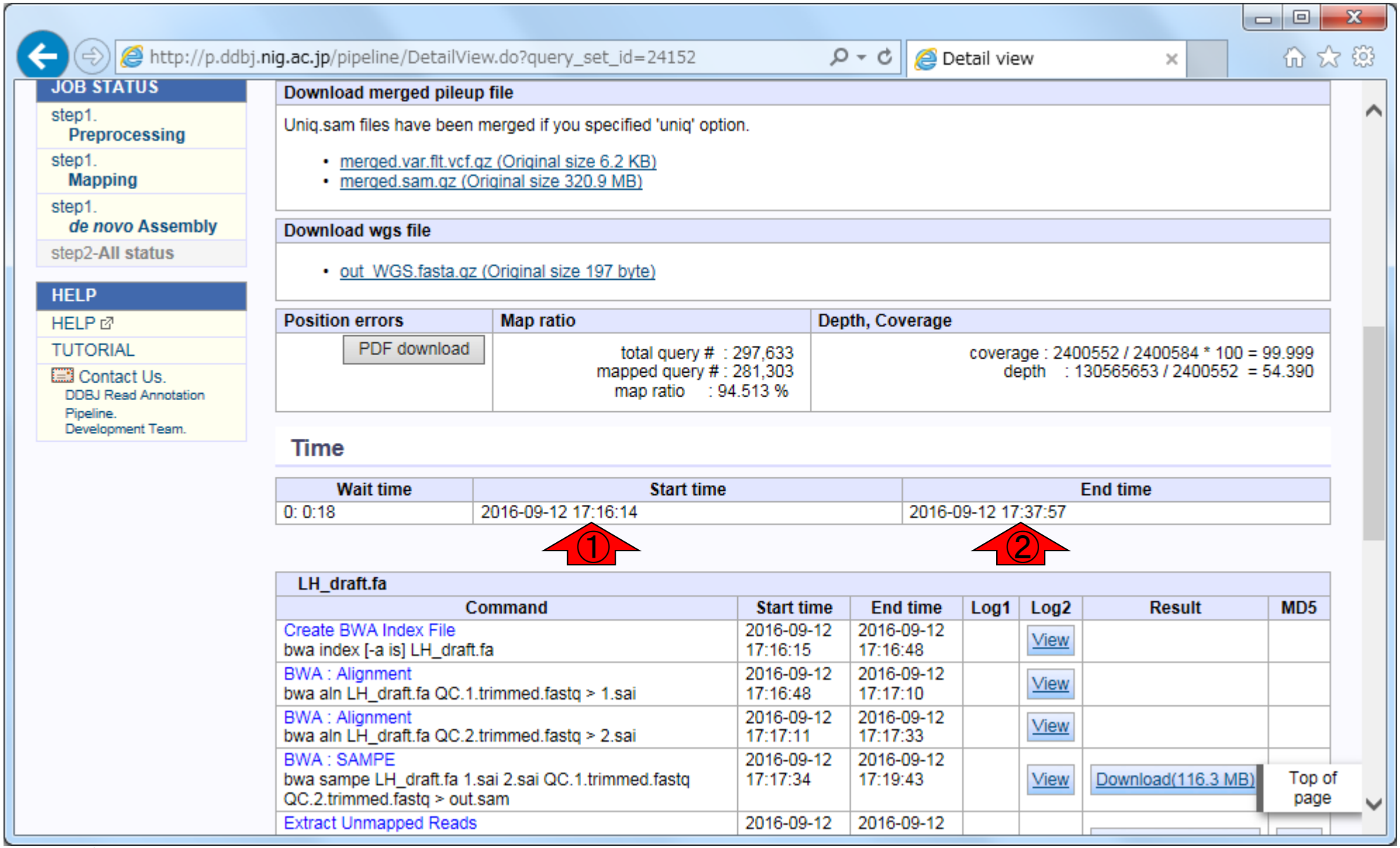
W15-2：結果の要

①Depth, Coverageのところの解説で、次は②depth。③マップされたリードで覆われている領域(2,400,552 bp)が平均してどれだけの厚み(depth)でマップされているかを示す。④130,565,653は、マップされたリードの総塩基数。リード長が251 bp、paired-endなので×2、マップされたリード数が281,303なので、251×2×281,303 = 141,214,106。この値は、リード長が全て251 bpであった場合であり、実際にはFaQCsでアダプタートリムやクオリティフィルタリングがかけられているので、ちょっと小さ目の値になるのは妥当



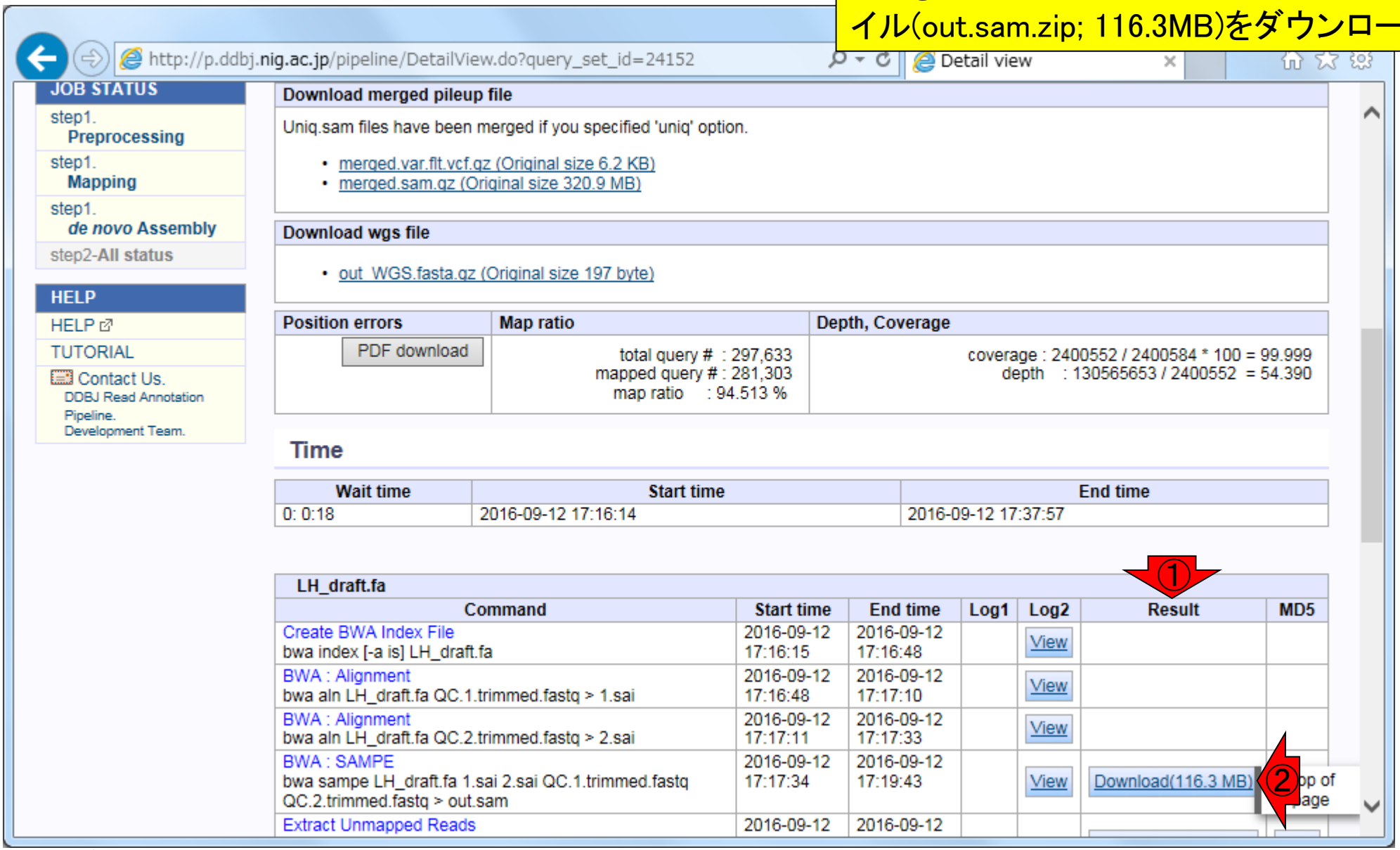
計算時間は「②－①」で、約20分

W15-2：結果の要約



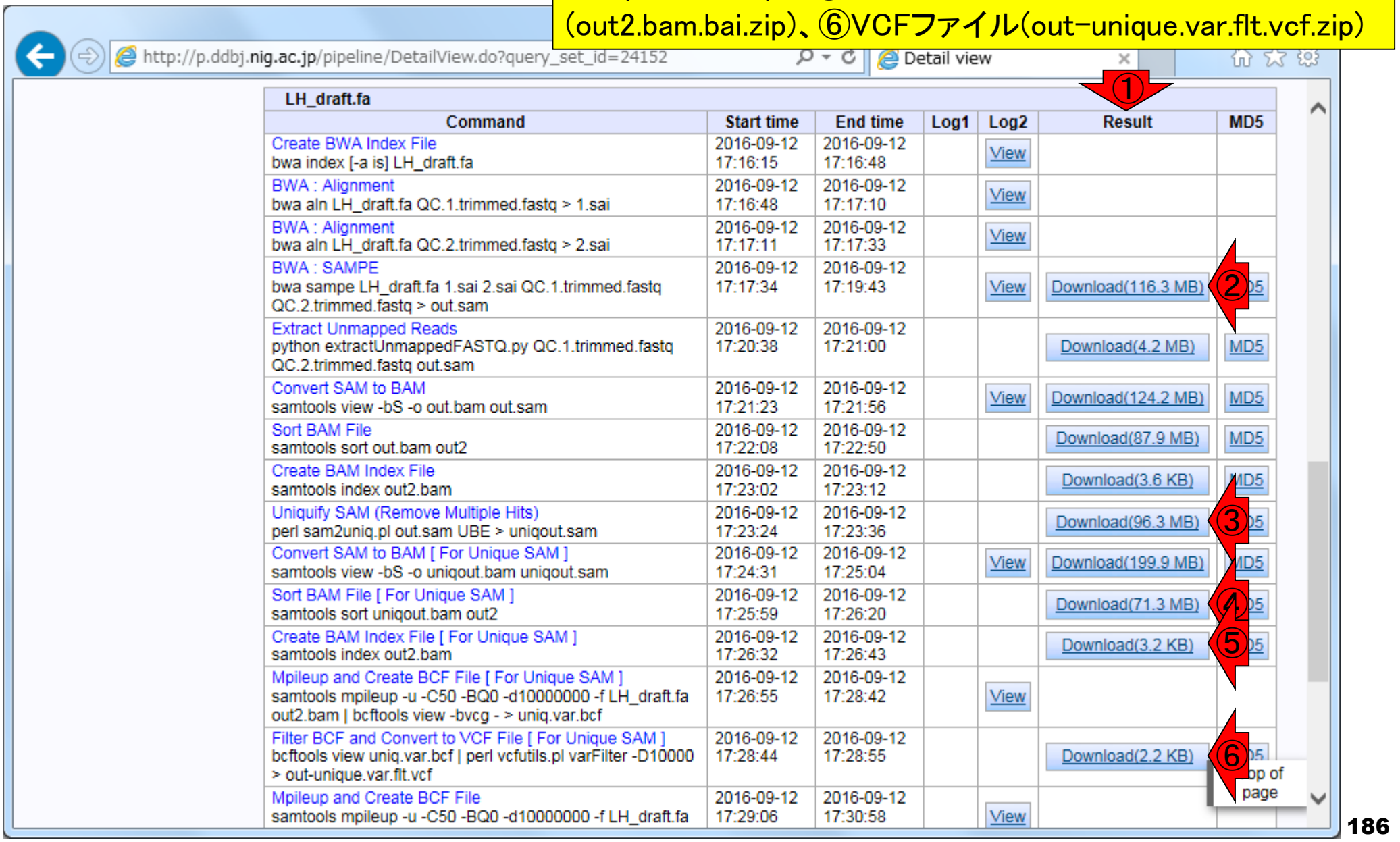
W15-3：ダウンロード

①DDBJ Pipelineでは、いくつかのBWA実行結果ファイルをダウンロード可能。まずは、②ユニーク化処理**前**のSAM形式ファイル(out.sam.zip; 116.3MB)をダウンロード



W15-3：ダウンロ

③ユニーク化処理後のSAM形式ファイル(uniqout.sam.zip)、④ユニーク化処理後のソート済みのBAM形式ファイル(uniqout.bam.zip)、⑤BAMインデックスファイル(out2.bam.bai.zip)、⑥VCFファイル(out-unique.var.flt.vcf.zip)

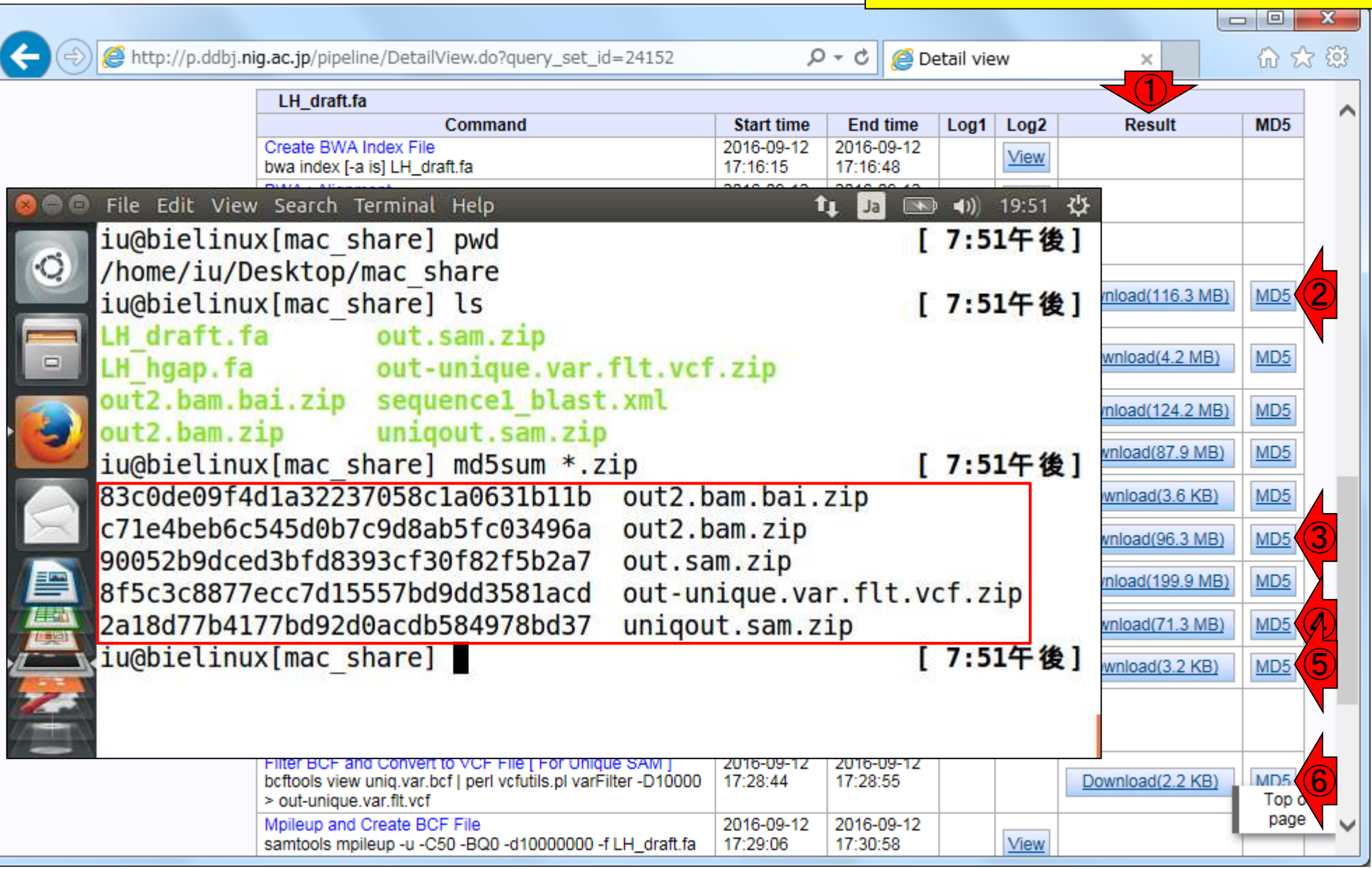


LH_draft.fa

Command	Start time	End time	Log1	Log2	Result	MD5
Create BWA Index File bwa index [-a is] LH_draft.fa	2016-09-12 17:16:15	2016-09-12 17:16:48		View		
BWA : Alignment bwa aln LH_draft.fa QC.1.trimmed.fastq > 1.sai	2016-09-12 17:16:48	2016-09-12 17:17:10		View		
BWA : Alignment bwa aln LH_draft.fa QC.2.trimmed.fastq > 2.sai	2016-09-12 17:17:11	2016-09-12 17:17:33		View		
BWA : SAMPE bwa sampe LH_draft.fa 1.sai 2.sai QC.1.trimmed.fastq QC.2.trimmed.fastq > out.sam	2016-09-12 17:17:34	2016-09-12 17:19:43		View	Download(116.3 MB)	05
Extract Unmapped Reads python extractUnmappedFASTQ.py QC.1.trimmed.fastq QC.2.trimmed.fastq out.sam	2016-09-12 17:20:38	2016-09-12 17:21:00			Download(4.2 MB)	MD5
Convert SAM to BAM samtools view -bS -o out.bam out.sam	2016-09-12 17:21:23	2016-09-12 17:21:56		View	Download(124.2 MB)	MD5
Sort BAM File samtools sort out.bam out2	2016-09-12 17:22:08	2016-09-12 17:22:50			Download(87.9 MB)	MD5
Create BAM Index File samtools index out2.bam	2016-09-12 17:23:02	2016-09-12 17:23:12			Download(3.6 KB)	MD5
Uniquify SAM (Remove Multiple Hits) perl sam2uniq.pl out.sam UBE > uniqout.sam	2016-09-12 17:23:24	2016-09-12 17:23:36			Download(96.3 MB)	05
Convert SAM to BAM [For Unique SAM] samtools view -bS -o uniqout.bam uniqout.sam	2016-09-12 17:24:31	2016-09-12 17:25:04		View	Download(199.9 MB)	MD5
Sort BAM File [For Unique SAM] samtools sort uniqout.bam out2	2016-09-12 17:25:59	2016-09-12 17:26:20			Download(71.3 MB)	05
Create BAM Index File [For Unique SAM] samtools index out2.bam	2016-09-12 17:26:32	2016-09-12 17:26:43			Download(3.2 KB)	05
Mpileup and Create BCF File [For Unique SAM] samtools mpileup -u -C50 -BQ0 -d10000000 -f LH_draft.fa out2.bam | bcftools view -bvcg - > uniq.var.bcf	2016-09-12 17:26:55	2016-09-12 17:28:42		View		
Filter BCF and Convert to VCF File [For Unique SAM] bcftools view uniq.var.bcf | perl vcfutils.pl varFilter -D10000 > out-unique.var.flt.vcf	2016-09-12 17:28:44	2016-09-12 17:28:55			Download(2.2 KB)	05
Mpileup and Create BCF File samtools mpileup -u -C50 -BQ0 -d10000000 -f LH_draft.fa	2016-09-12 17:29:06	2016-09-12 17:30:58		View		

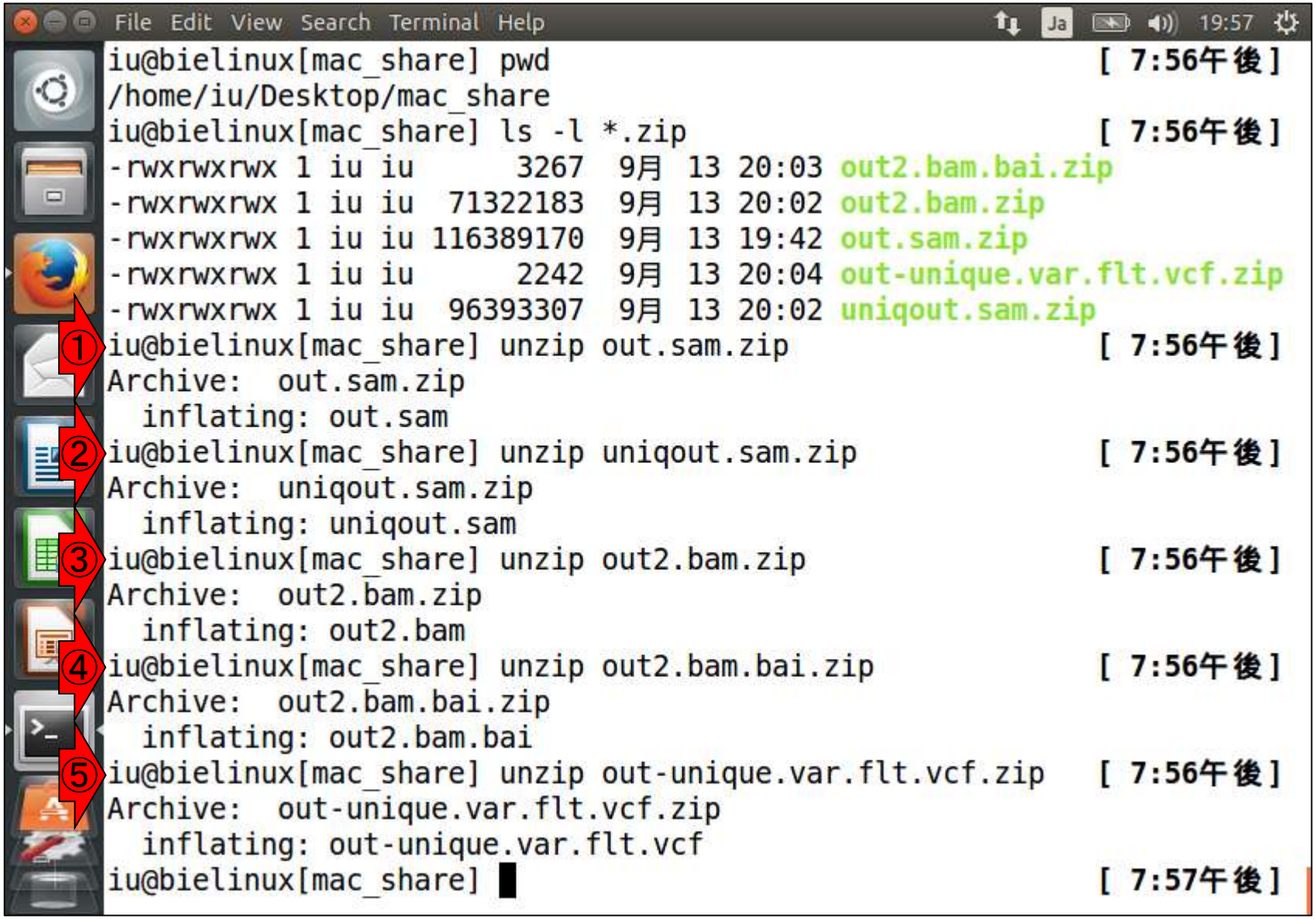
W15-3：ダウンロード

md5sumコマンドで確認。②から⑥で提示されている文字列と同じであればOK。第3回W12、第7回W9-3にもあり



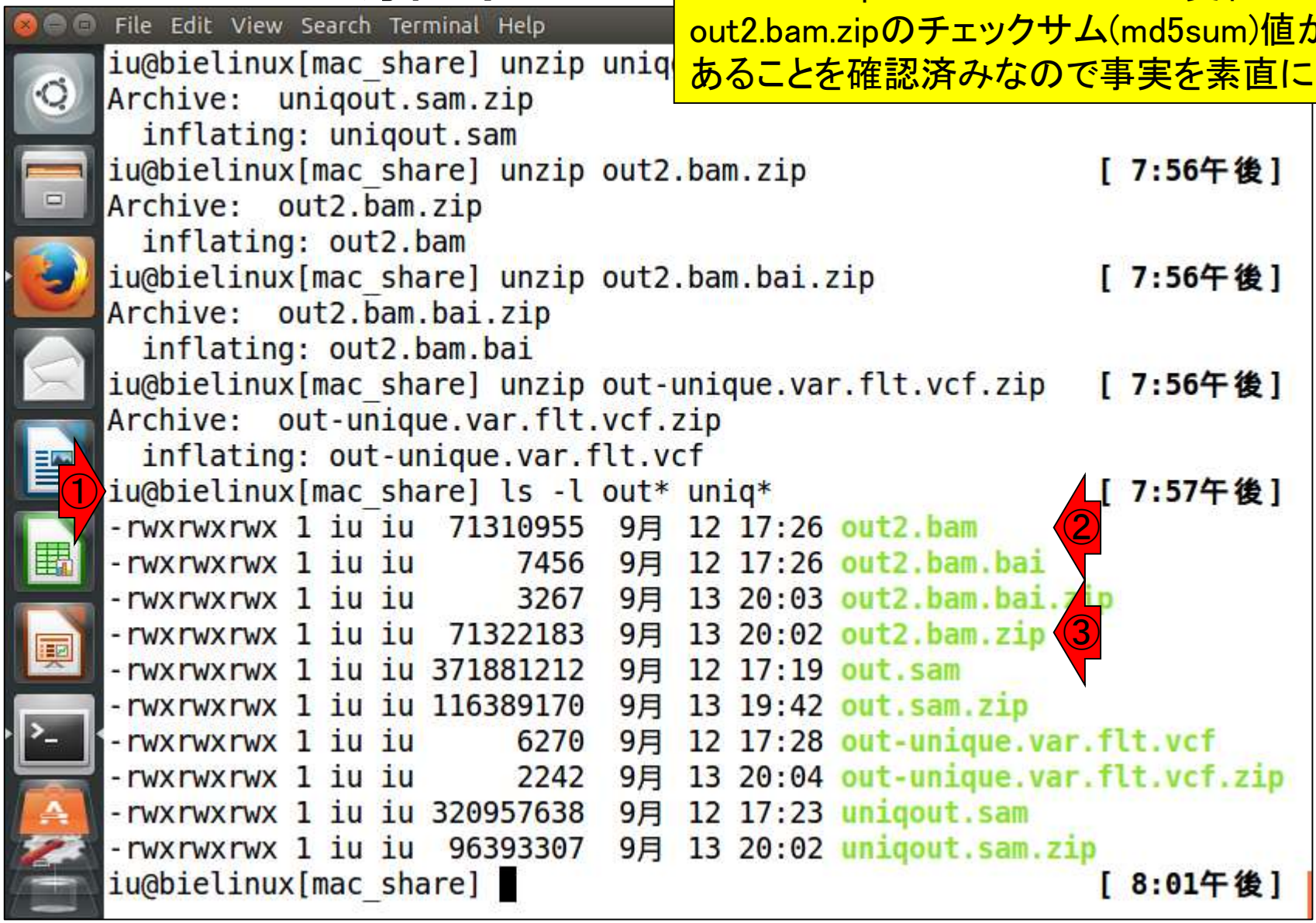
ダウンロードしたzipファイルの解凍

W16-1：解凍



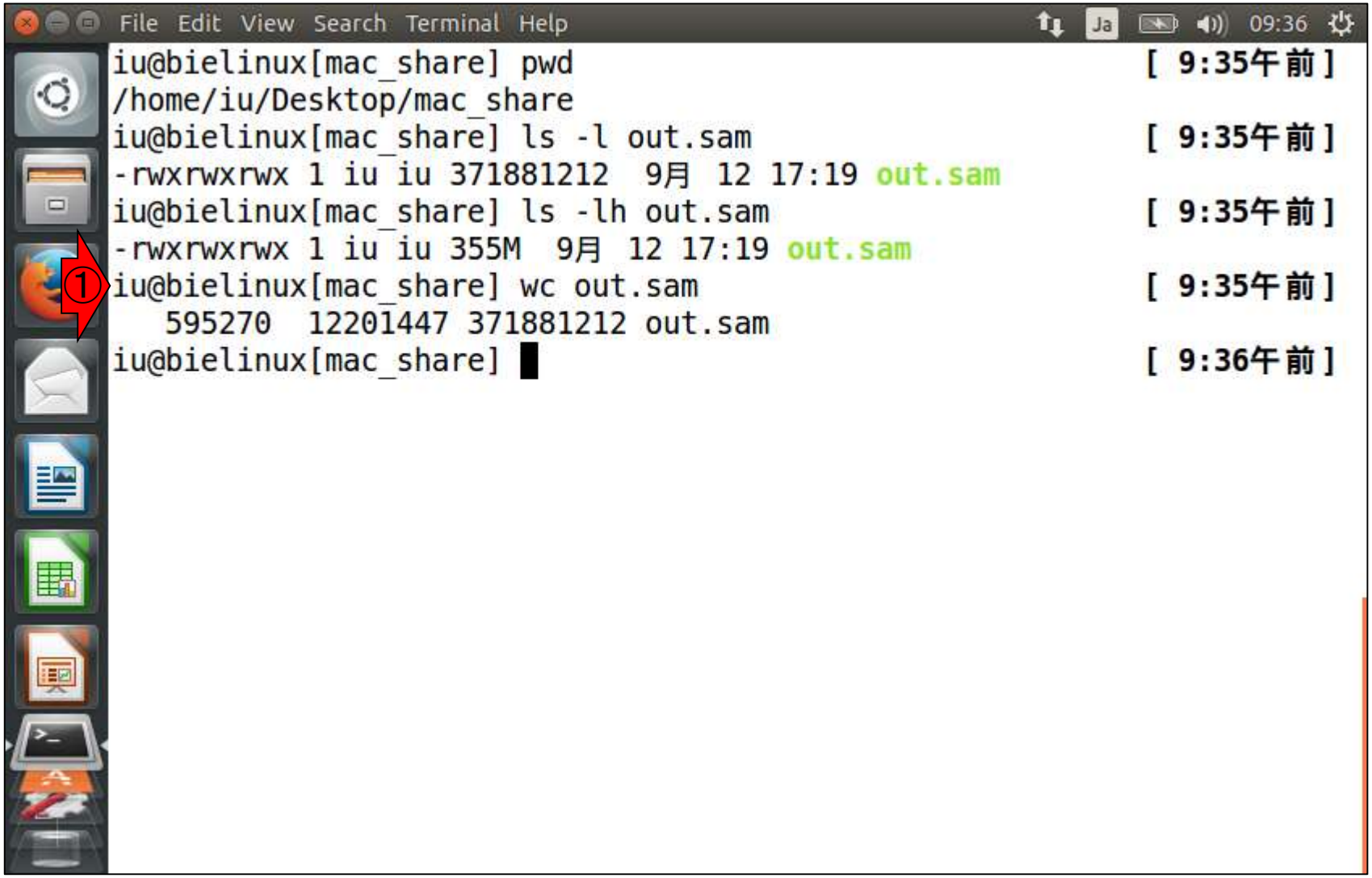
W16-1：解凍

①解凍後にファイルサイズを確認。全体的に圧縮後に2-3倍程度になっている。②out2.bamのみ、③解凍前のout2.bam.zipとサイズがほとんど変わっていないが、out2.bam.zipのチェックサム(md5sum)値が提供元と同じであることを確認済みなので事実を素直に受け止めるのみ



W16-2：out.sam

①out.samファイル(371,881,212 bytes; 355MB)は、595,270行

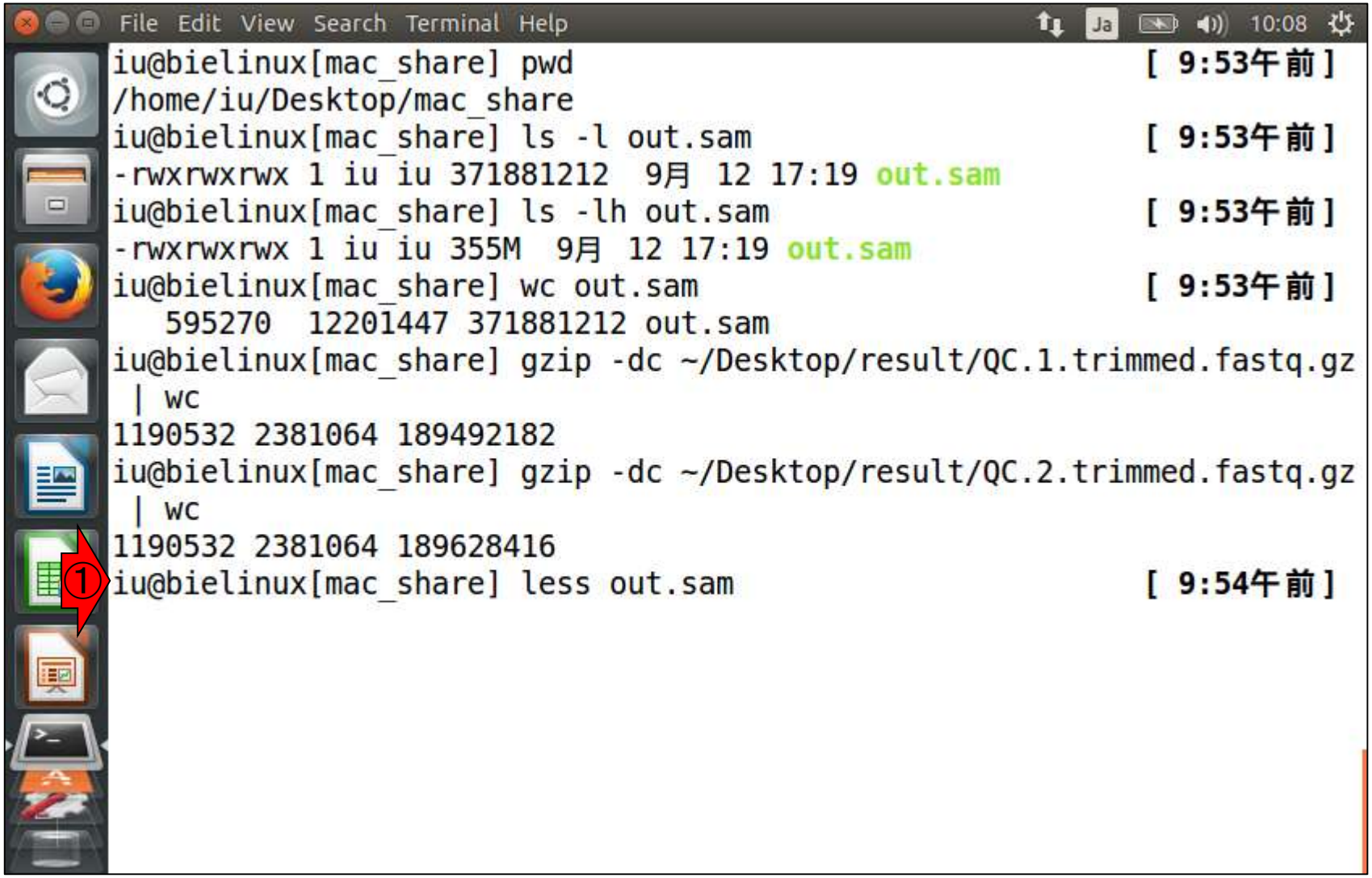


W16-3：マップする側

マップする側のファイル(QC.*.trimmed.fastq.gz)の行数をおさらい。1,190,532行。FASTQ形式なので1,190,532 / 4 = 297,633リード。①②gzipのdオプションが解凍、cオプションが元ファイルを残す指定

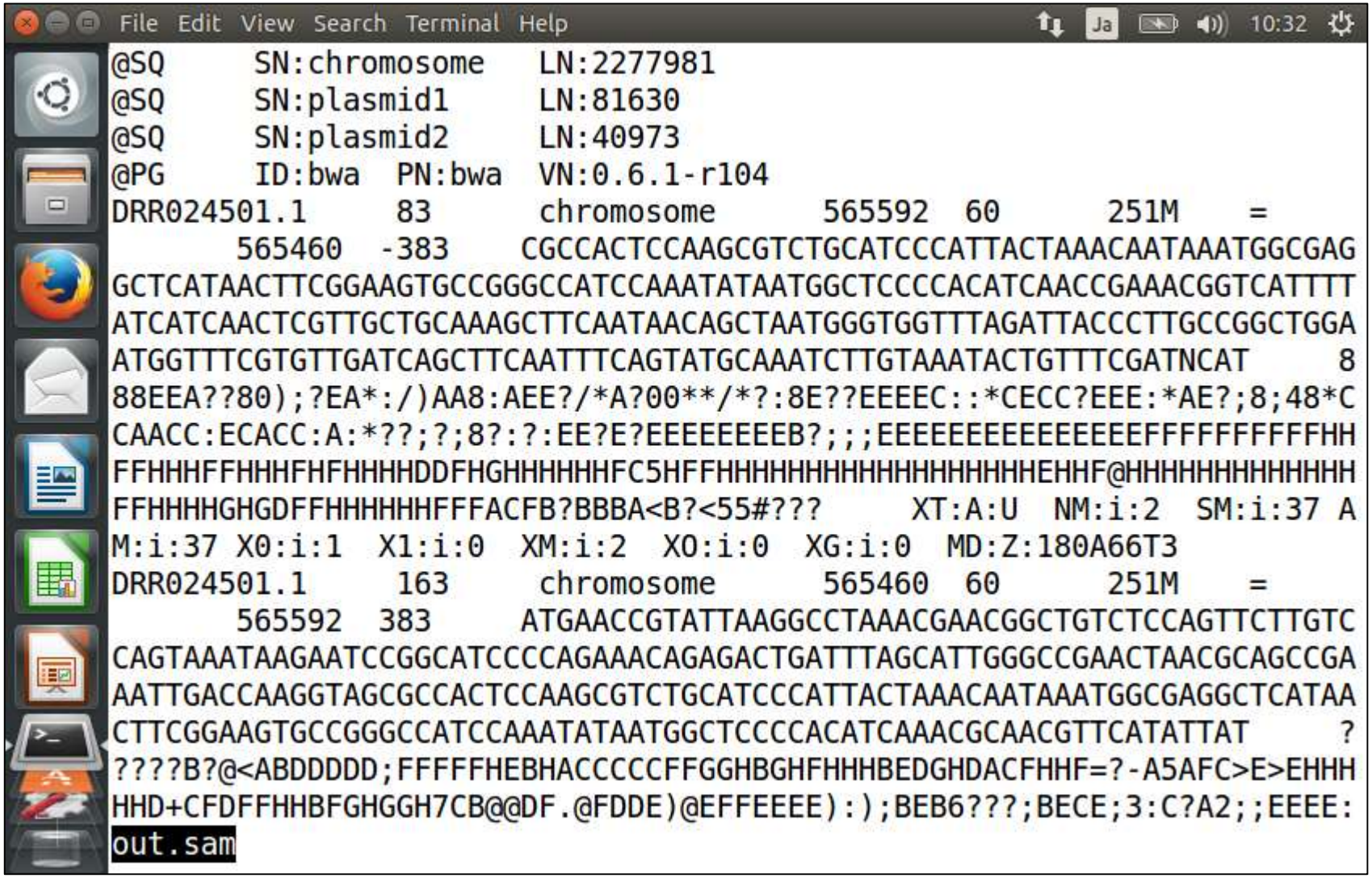
①lessで眺める

W16-4：less out.sam



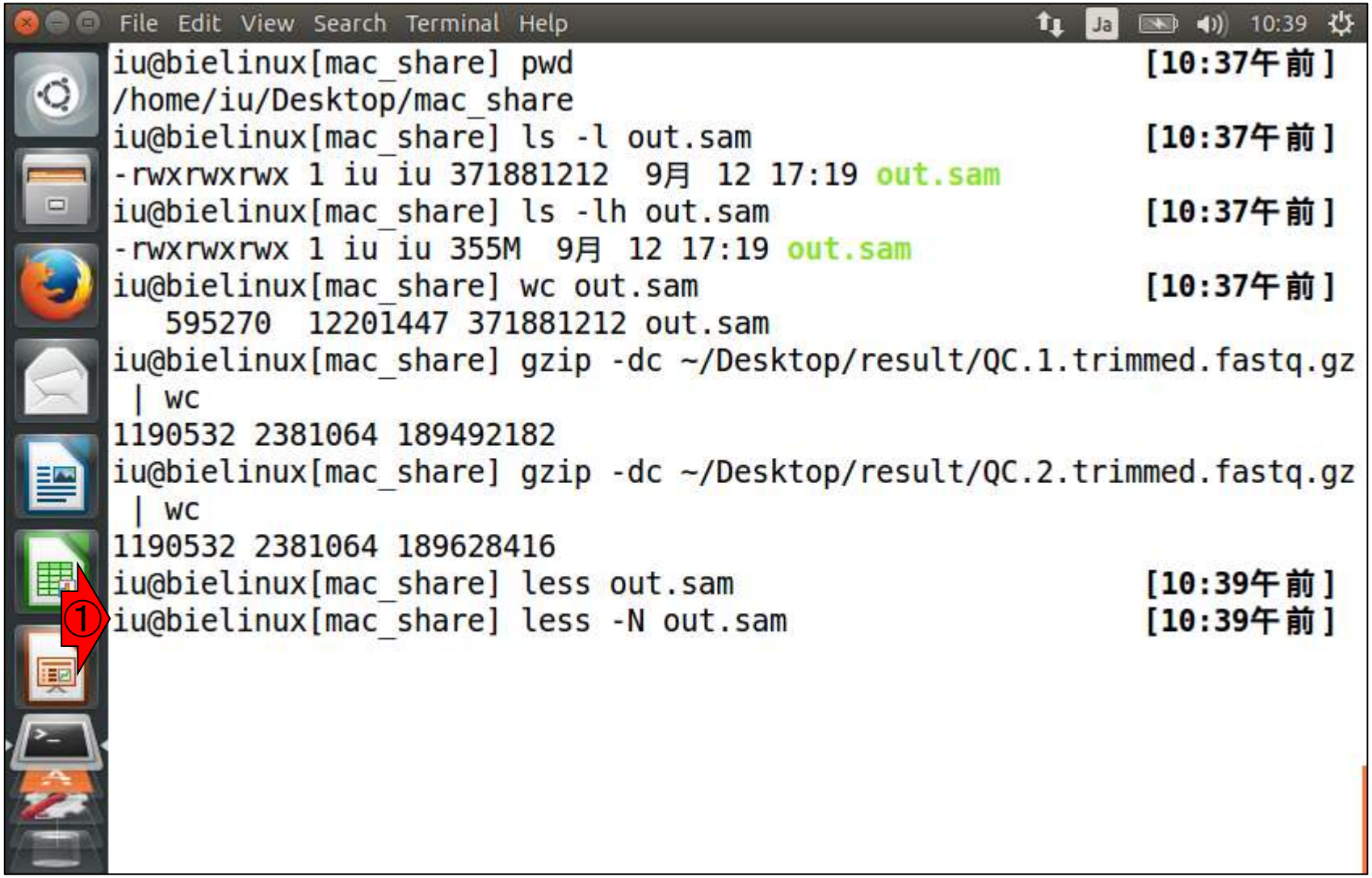
W16-4：lessで眺める

①lessでout.samを眺める。非常に見づらいので、行番号を表示したい。一旦qで抜ける



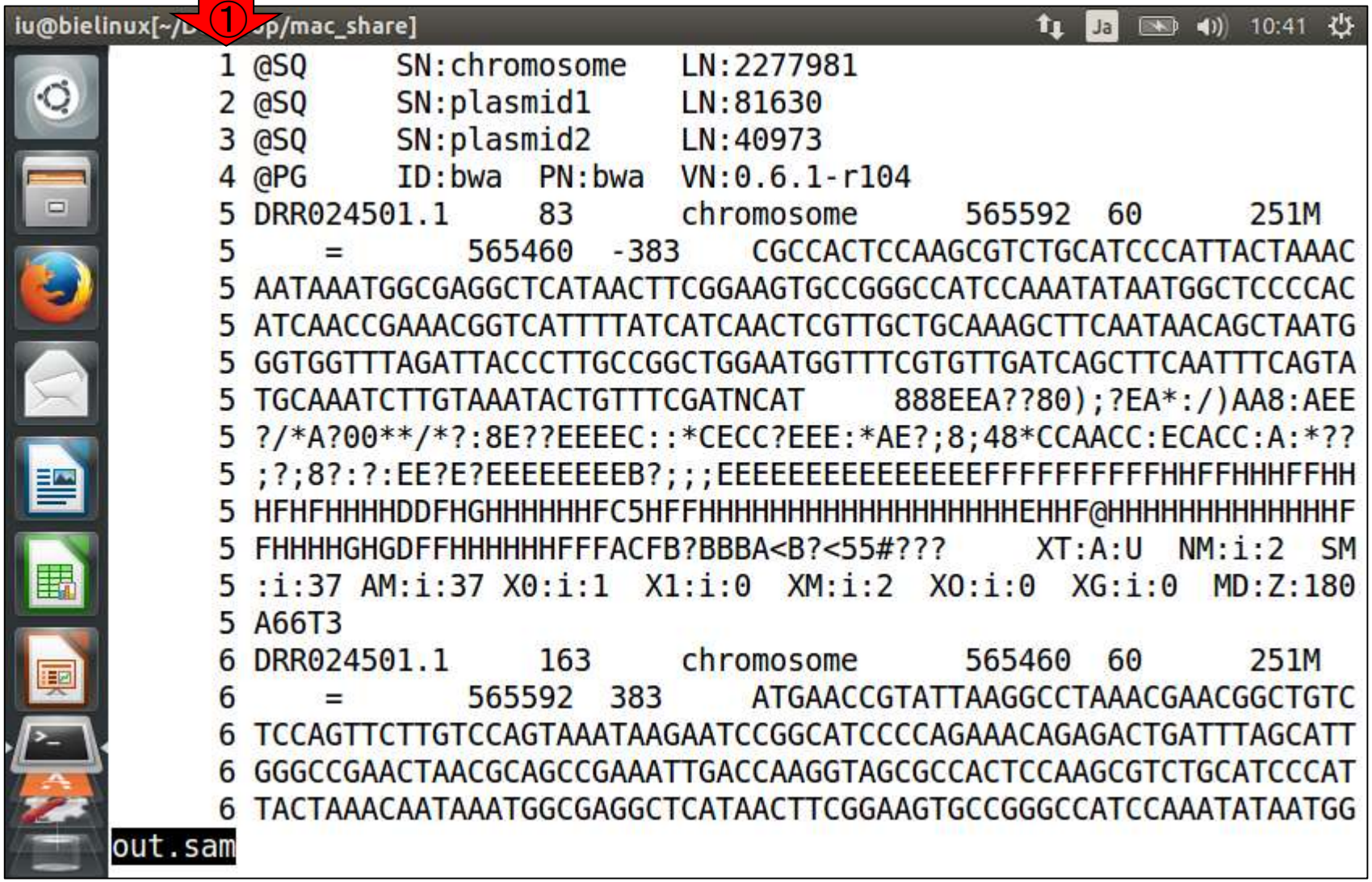
W16-4：lessで眺める

行番号も表示させたいときは、①
less実行時にNオプションをつける



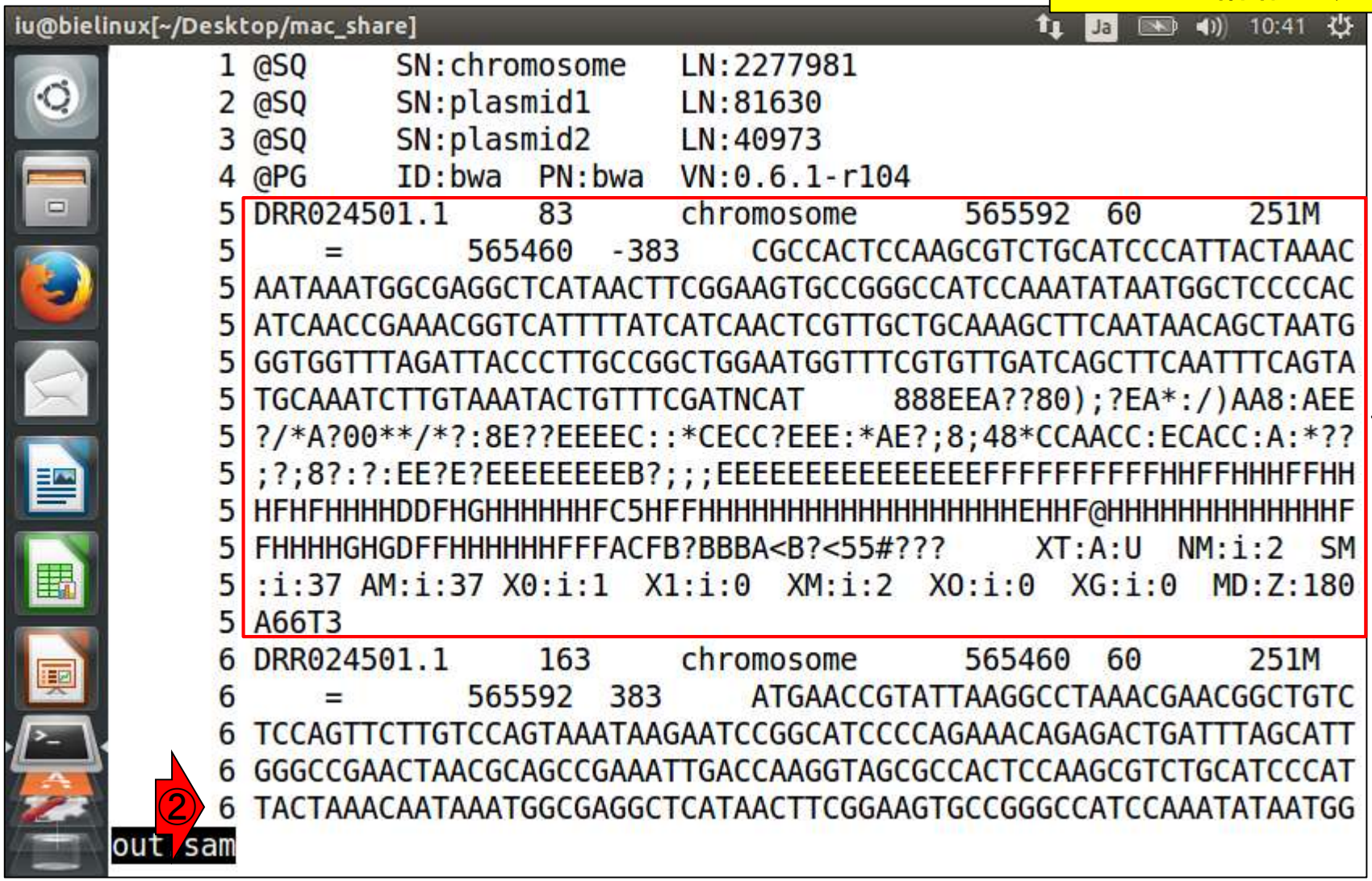
こんな感じになります。①が行番号情報

W16-4：lessで眺める



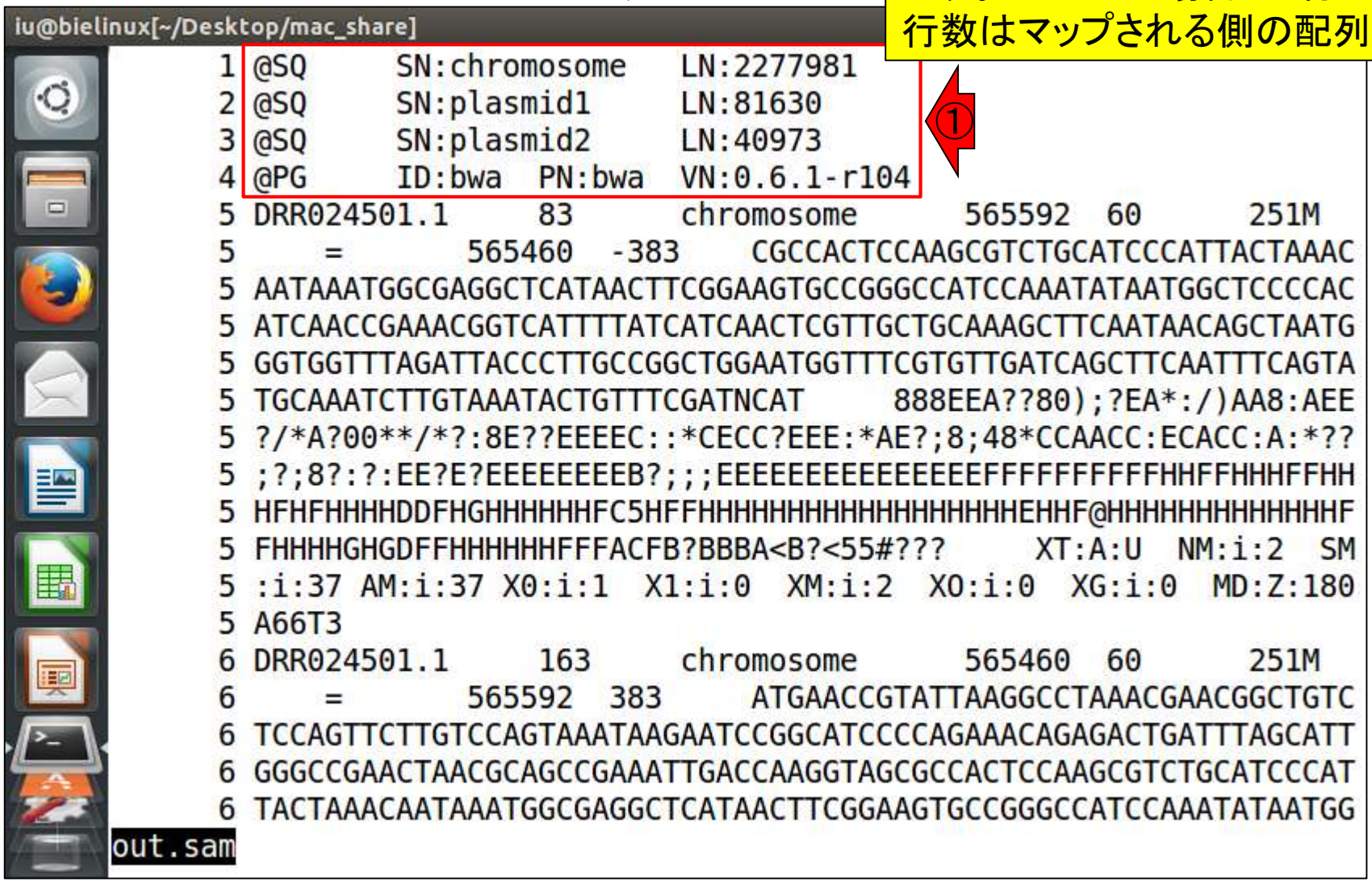
W16-4：lessで眺める

例えば①の赤枠内が5行目の情報。この画面上では、おそらく6行目の途中までの情報が表示されています



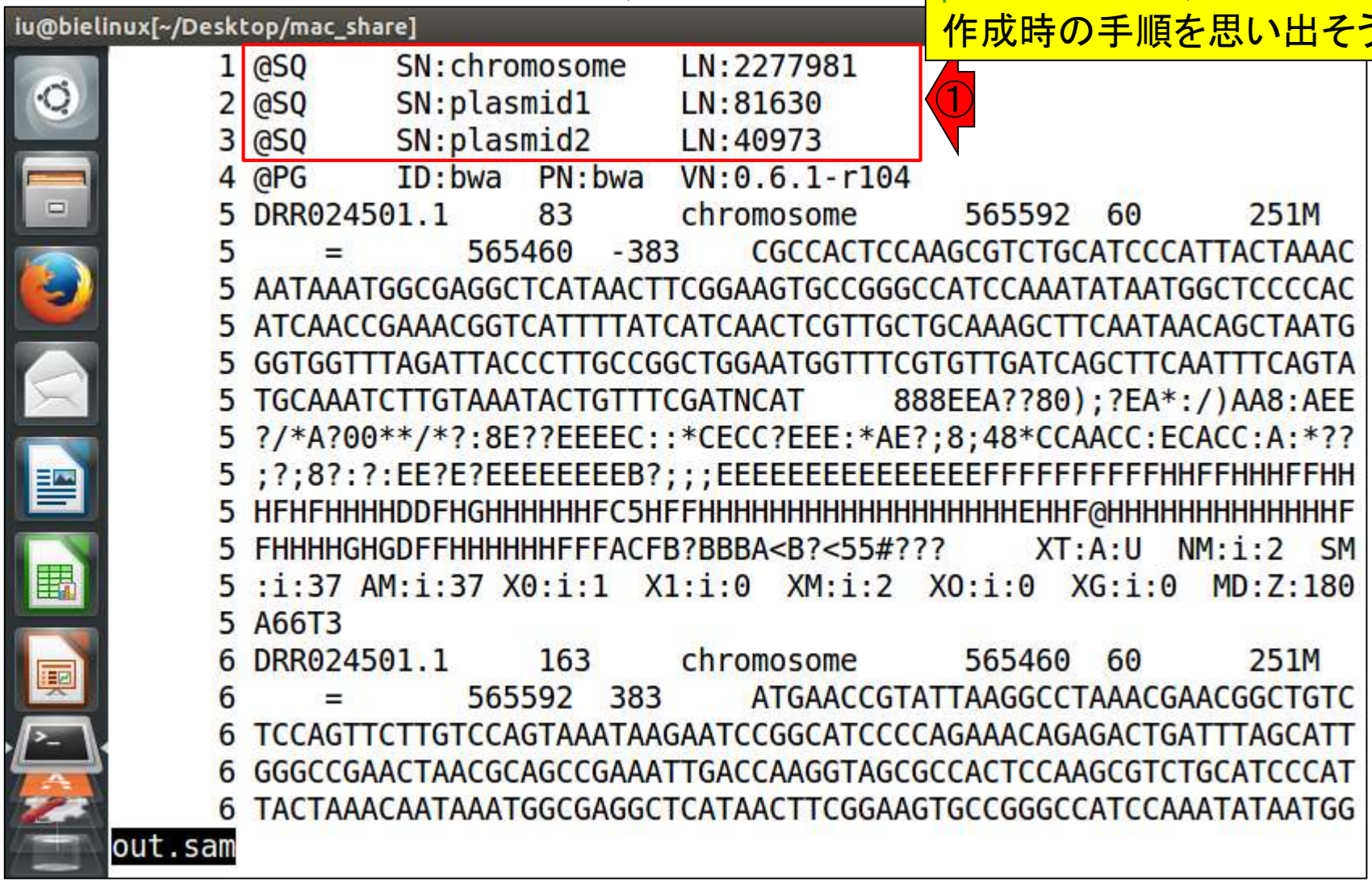
W16-4：lessで眺める

今はout.samを眺めることで、SAM形式を学んでいます。①@からはじまる行は、ヘッダー行です。out.samの場合は4行ですが、ヘッダーの行数はマップされる側の配列数に依存します



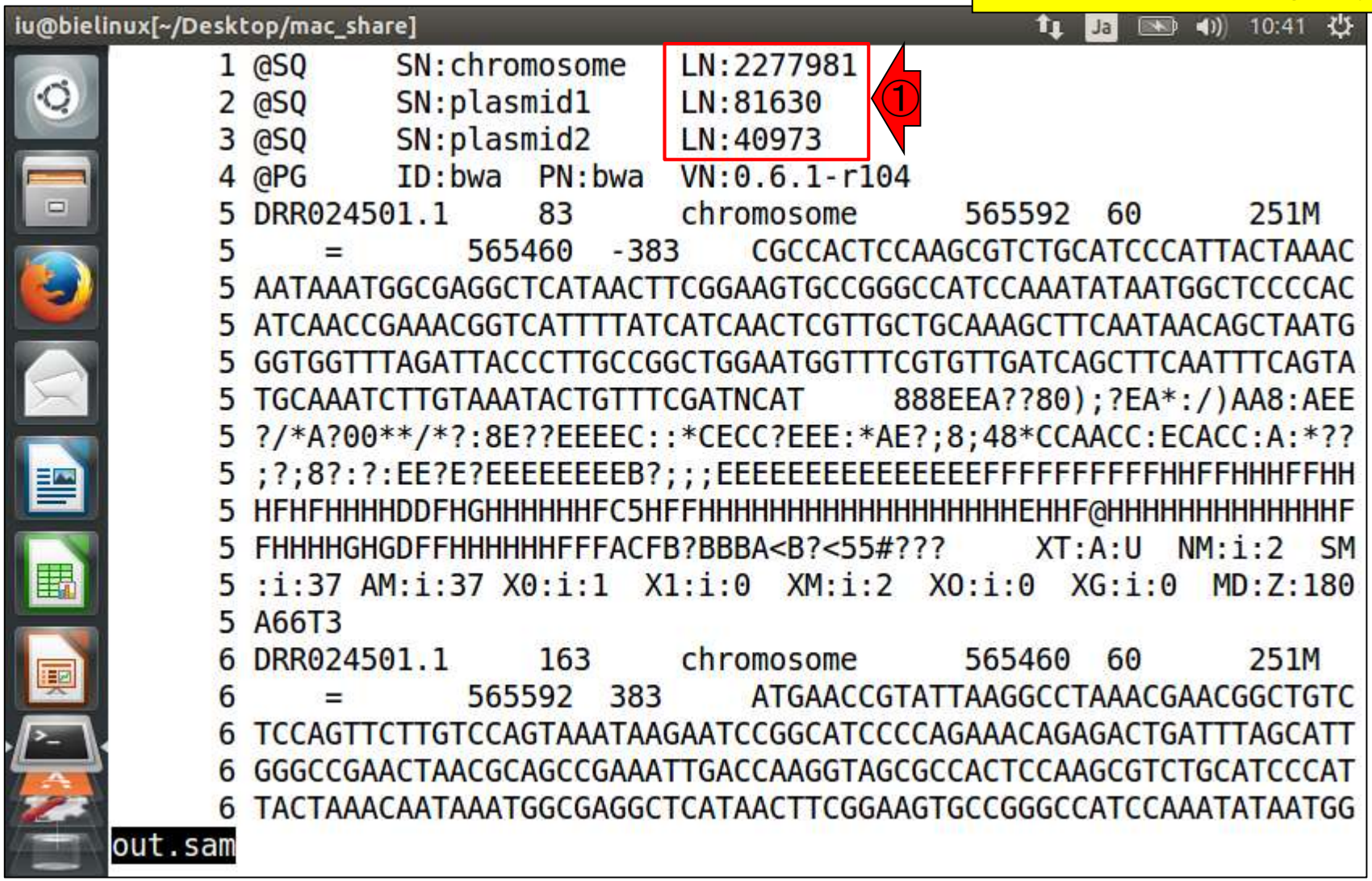
W16-4：lessで眺める

配列数依存の意味は、ここを見ればわかる。description部分をchromosome, plasmid1, plasmid2としたリファレンス配列(LH_draft.fa)作成時の手順を思い出そう(W11-2)



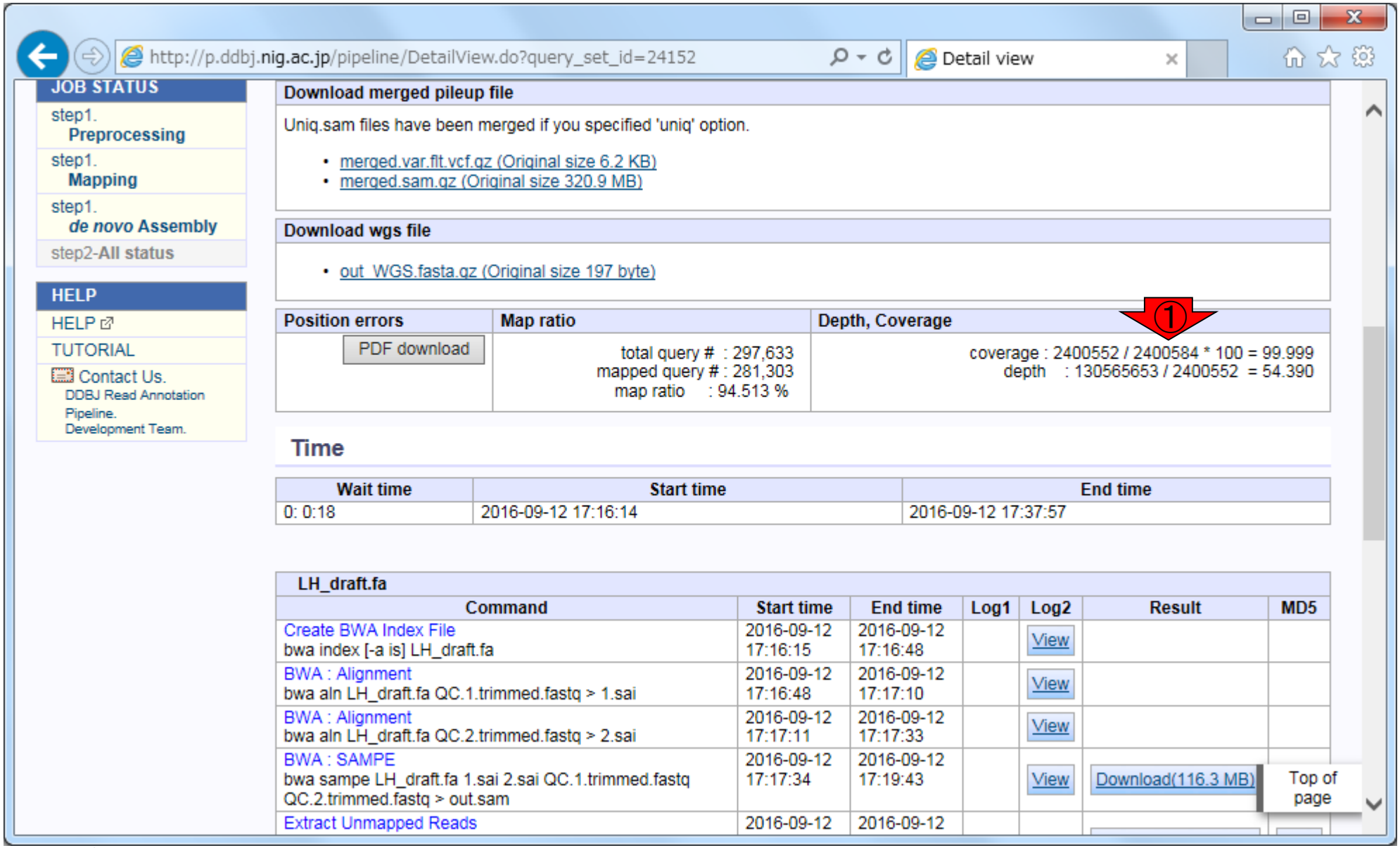
W16-4：lessで眺める

ここの数値は配列ごとの塩基数です。
2,277,981 + 81,630 + 40,973 = 2,400,584
がゲノムサイズに相当。この数値は…



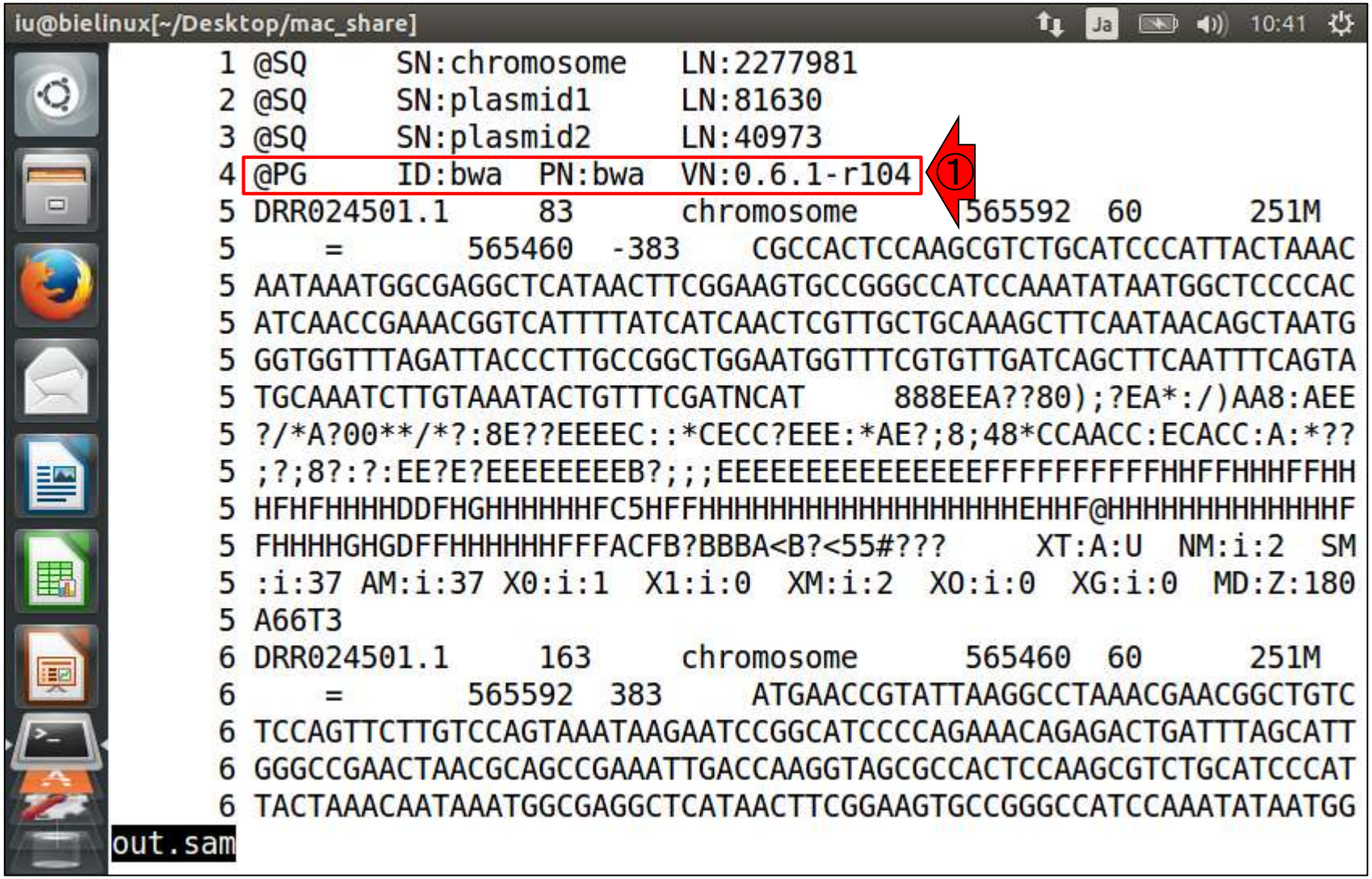
W16-4：lessで眺める

DDBJ Pipeline実行結果画面上の、
①と同じことからも納得できるでしょう



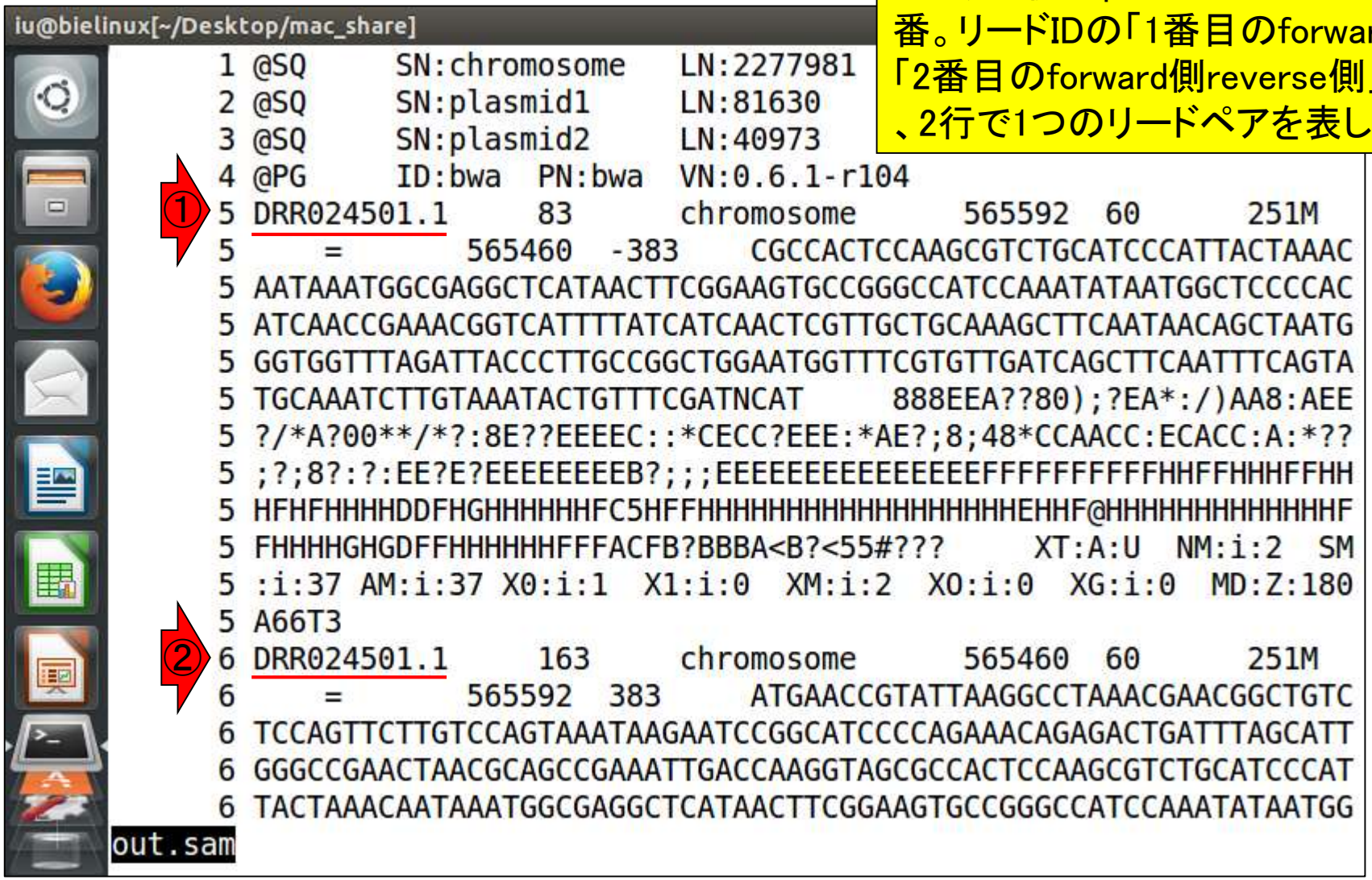
W16-4：lessで眺める

①はマッピングに用いたプログラム、およびバージョン情報が記載されています



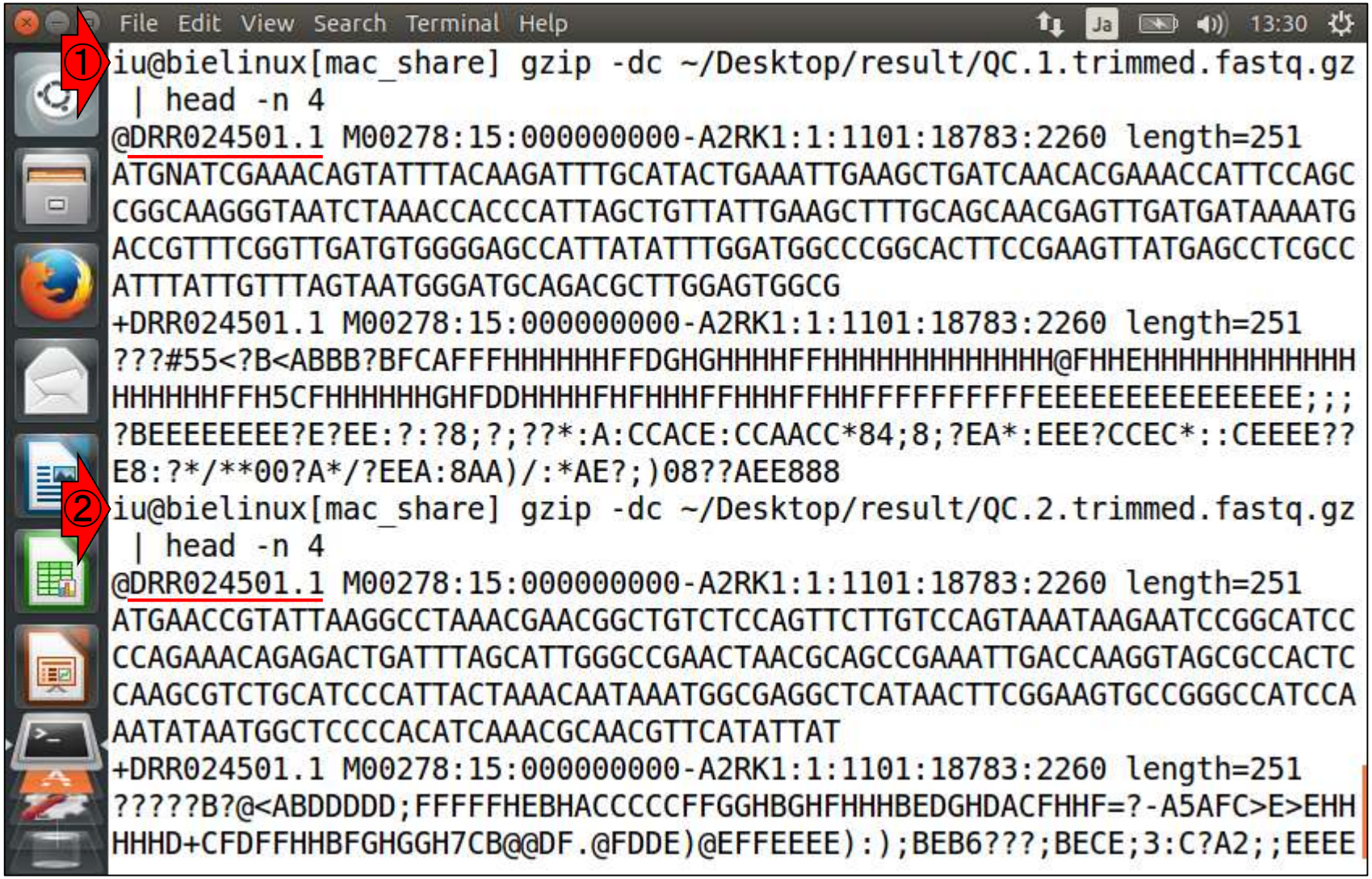
W16-5:5行目以降

①5行目の最初、②6行目の最初をみると、同じリードIDになっていることがわかる。並びは、マップする側のpaired-endファイルのリードの順番。リードIDの「1番目のforward側reverse側」、「2番目のforward側reverse側」、のような感じで、2行で1つのリードペアを表している



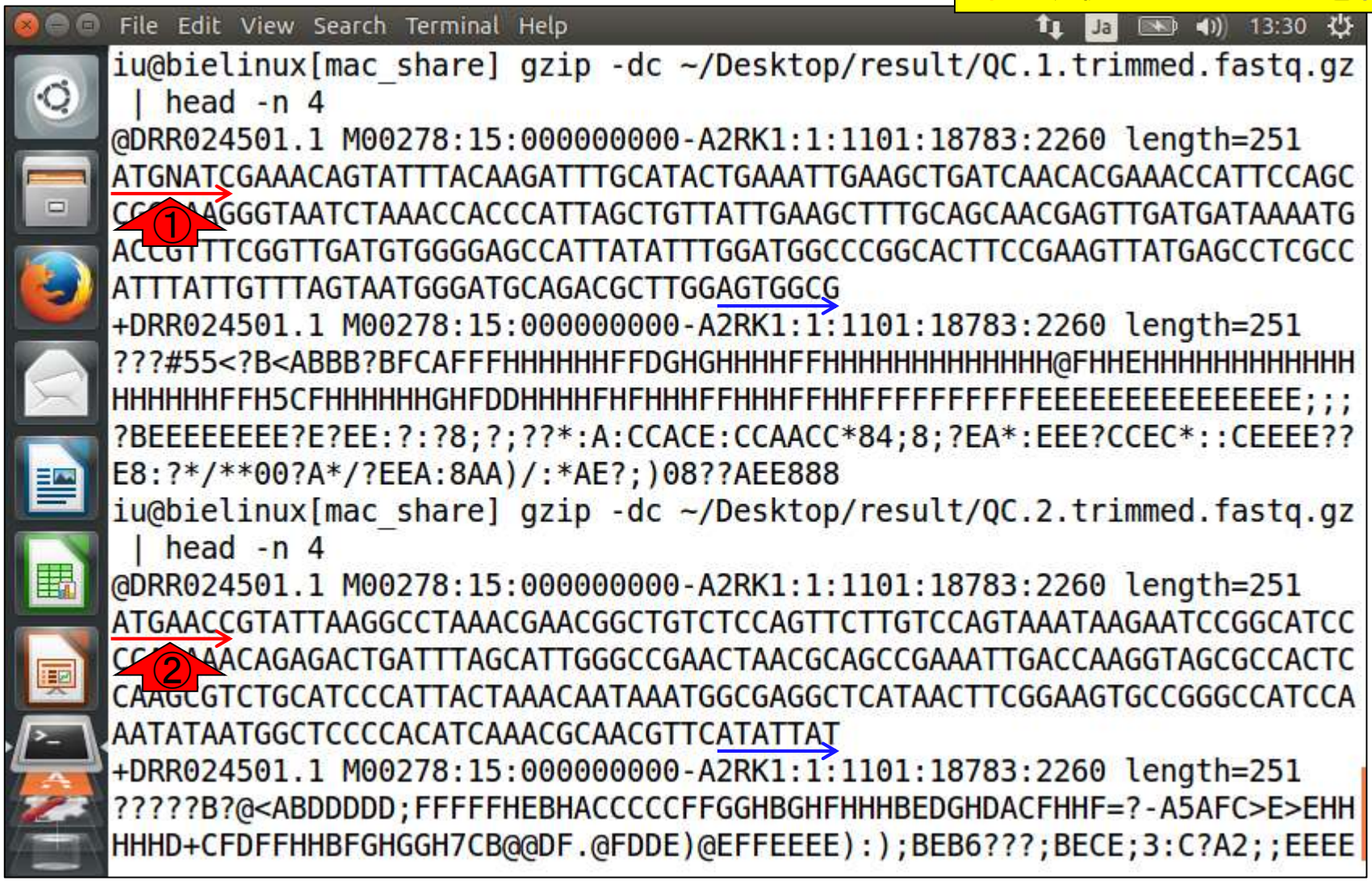
W16-5：5行目以降

リードIDの「1番目の①forward側②reverse側」を表示。確かに同じリードID



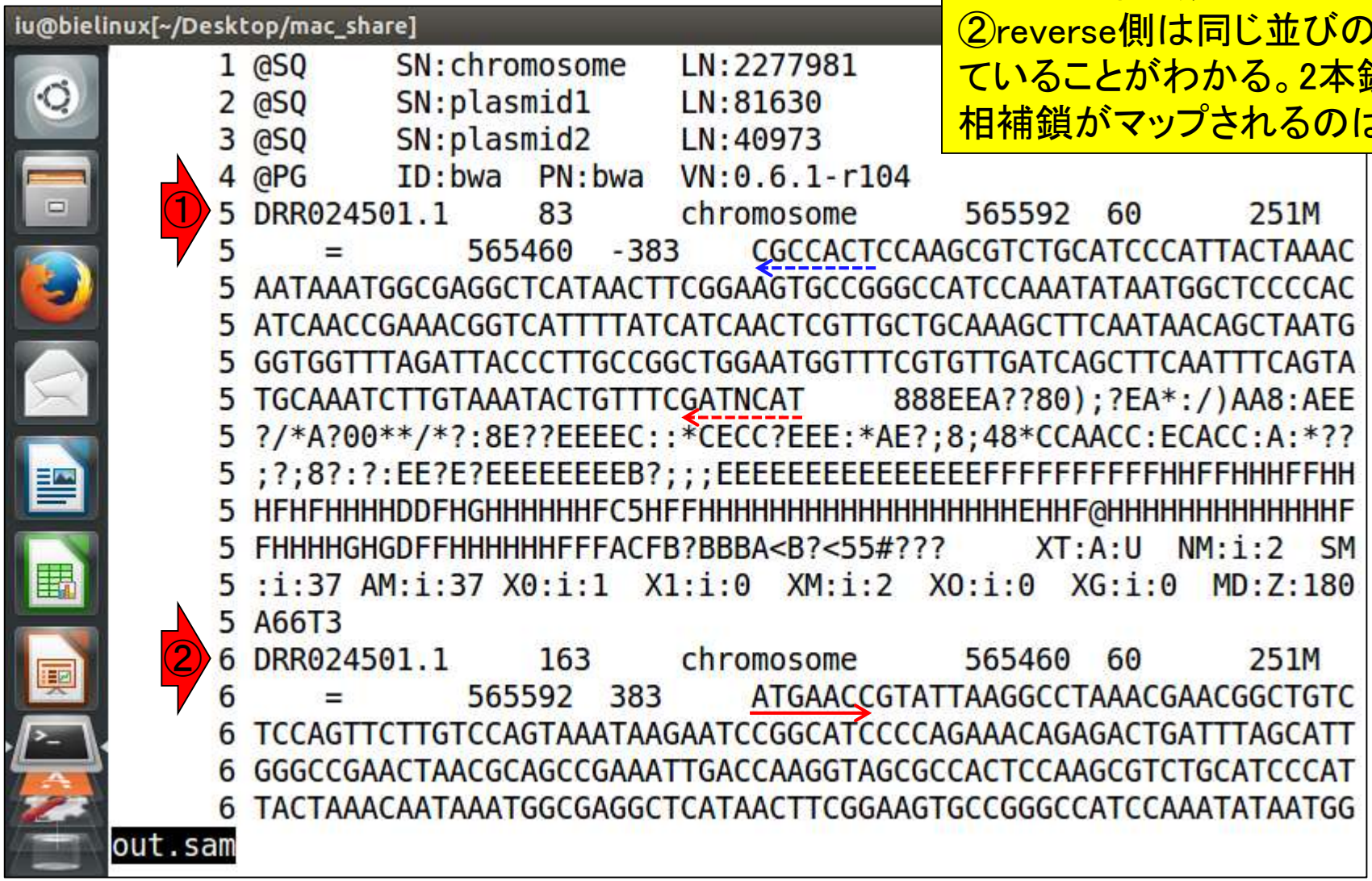
W16-5：5行目以降

リードIDの「1番目の①forward側②reverse側」の塩基配列の最初の7塩基分を赤矢印で、最後の7塩基分を青矢印で示す



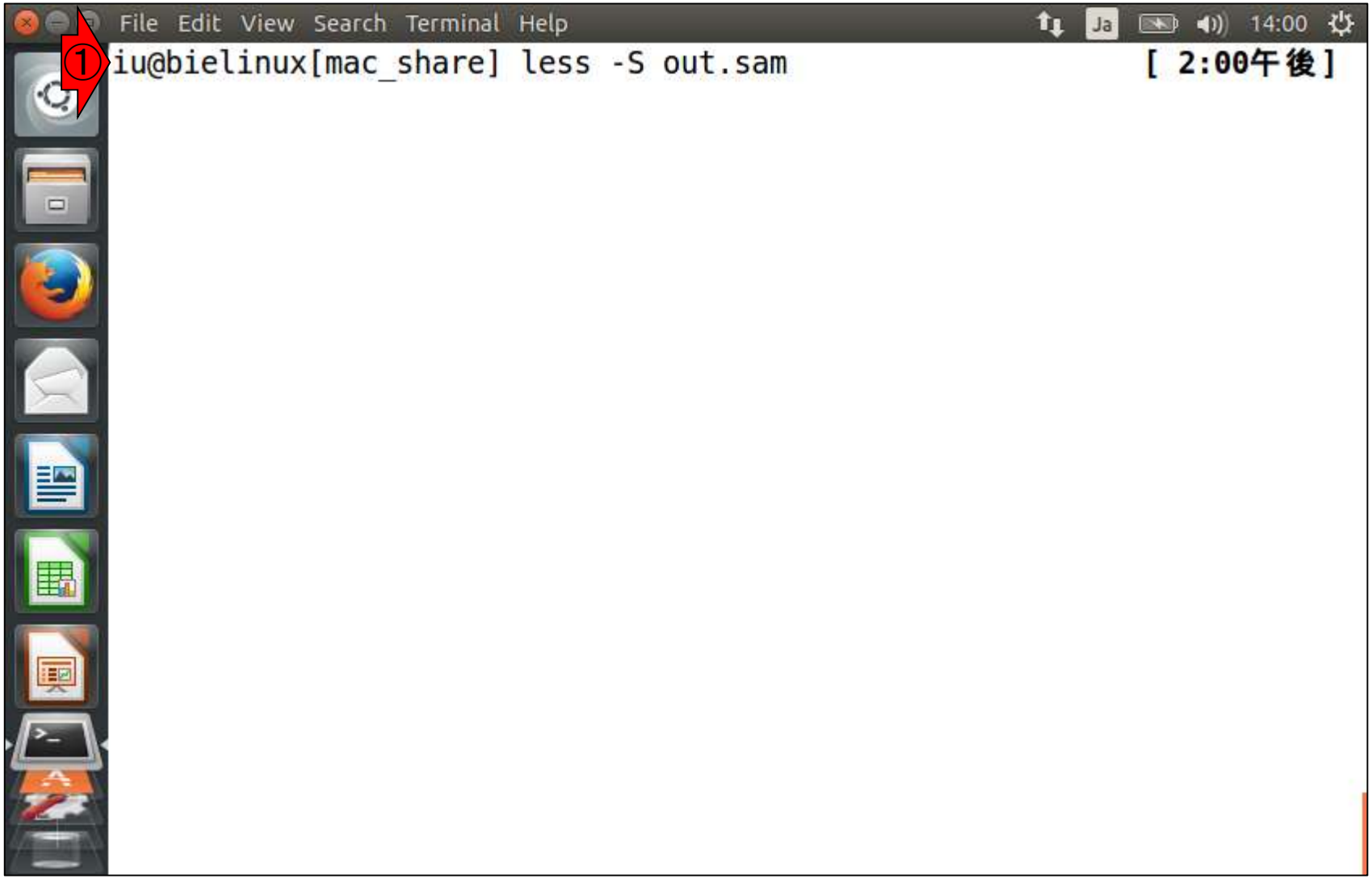
W16-5：5行目以降

「less -N out.sam」上で、さきほどの赤矢印に対応する部分を示す。①forward側のリードは逆相補鎖(reverse complement)が、②reverse側は同じ並びの塩基配列になっていることがわかる。2本鎖の場合は、逆相補鎖がマップされるのは普通なので妥当



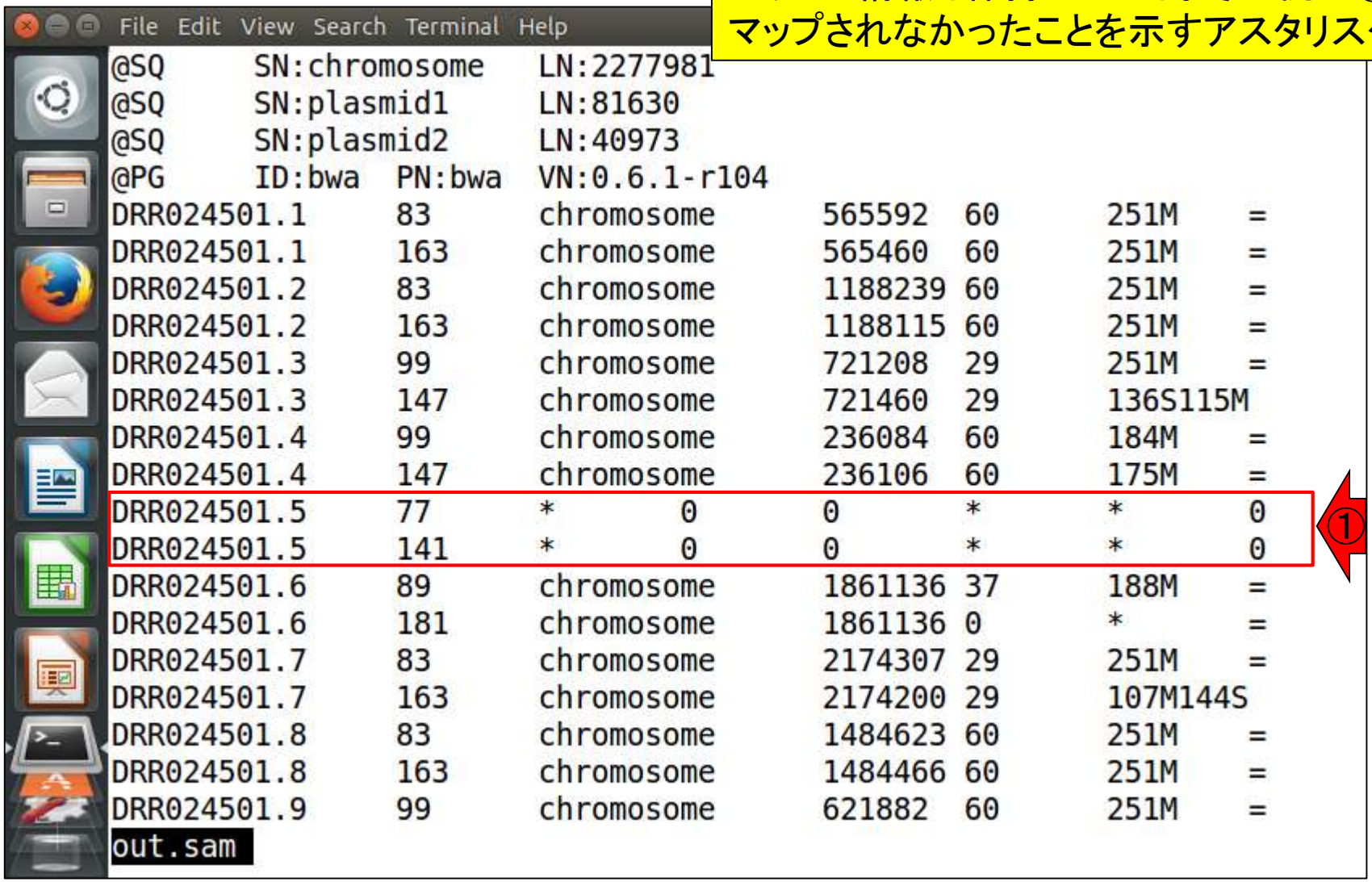
W16-6：less -S

less実行時にSオプションをつけると1行分を折り返さずに表示してくれます



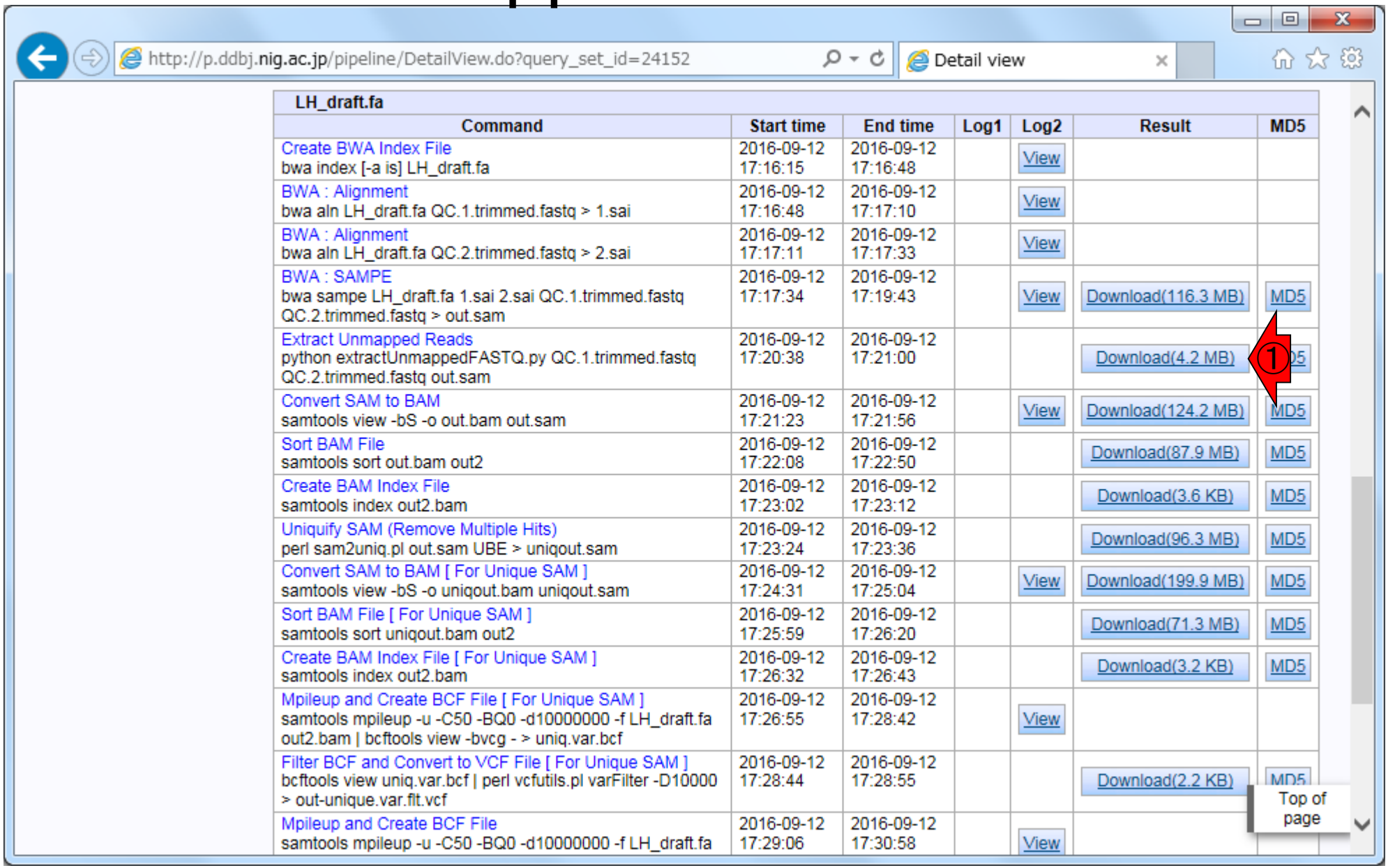
W16-6：less -S

こんな感じ。1行が長いファイルで、各行の最初のほうのみ眺めたい場合に便利。SAMファイルはマップされなかったリード情報も保持している。その例が①5番目のリード。マップされなかったことを示すアスタリスク(*)で認識可能



W16-7：unmapped reads

①マップされなかったリード情報のみ
FASTQ形式ファイルとして得たい場合



LH_draft.fa

Command	Start time	End time	Log1	Log2	Result	MD5
Create BWA Index File bwa index [-a is] LH_draft.fa	2016-09-12 17:16:15	2016-09-12 17:16:48		View		
BWA : Alignment bwa aln LH_draft.fa QC.1.trimmed.fastq > 1.sai	2016-09-12 17:16:48	2016-09-12 17:17:10		View		
BWA : Alignment bwa aln LH_draft.fa QC.2.trimmed.fastq > 2.sai	2016-09-12 17:17:11	2016-09-12 17:17:33		View		
BWA : SAMPE bwa sampe LH_draft.fa 1.sai 2.sai QC.1.trimmed.fastq QC.2.trimmed.fastq > out.sam	2016-09-12 17:17:34	2016-09-12 17:19:43		View	Download(116.3 MB)	MD5
Extract Unmapped Reads python extractUnmappedFASTQ.py QC.1.trimmed.fastq QC.2.trimmed.fastq out.sam	2016-09-12 17:20:38	2016-09-12 17:21:00			Download(4.2 MB)	MD5
Convert SAM to BAM samtools view -bS -o out.bam out.sam	2016-09-12 17:21:23	2016-09-12 17:21:56		View	Download(124.2 MB)	MD5
Sort BAM File samtools sort out.bam out2	2016-09-12 17:22:08	2016-09-12 17:22:50			Download(87.9 MB)	MD5
Create BAM Index File samtools index out2.bam	2016-09-12 17:23:02	2016-09-12 17:23:12			Download(3.6 KB)	MD5
Uniquify SAM (Remove Multiple Hits) perl sam2uniq.pl out.sam UBE > uniqout.sam	2016-09-12 17:23:24	2016-09-12 17:23:36			Download(96.3 MB)	MD5
Convert SAM to BAM [For Unique SAM] samtools view -bS -o uniqout.bam uniqout.sam	2016-09-12 17:24:31	2016-09-12 17:25:04		View	Download(199.9 MB)	MD5
Sort BAM File [For Unique SAM] samtools sort uniqout.bam out2	2016-09-12 17:25:59	2016-09-12 17:26:20			Download(71.3 MB)	MD5
Create BAM Index File [For Unique SAM] samtools index out2.bam	2016-09-12 17:26:32	2016-09-12 17:26:43			Download(3.2 KB)	MD5
Mpileup and Create BCF File [For Unique SAM] samtools mpileup -u -C50 -BQ0 -d10000000 -f LH_draft.fa out2.bam | bcftools view -bvcg - > uniq.var.bcf	2016-09-12 17:26:55	2016-09-12 17:28:42		View		
Filter BCF and Convert to VCF File [For Unique SAM] bcftools view uniq.var.bcf | perl vcfutils.pl varFilter -D10000 > out-unique.var.flt.vcf	2016-09-12 17:28:44	2016-09-12 17:28:55			Download(2.2 KB)	MD5
Mpileup and Create BCF File samtools mpileup -u -C50 -BQ0 -d10000000 -f LH_draft.fa	2016-09-12 17:29:06	2016-09-12 17:30:58		View		

W17-1：Excel

out.samをExcelで開いた状態。①QNAME（Query NAME）。リードIDのこと。②FLAG。フラグ情報（マップされたかどうかなどの情報だが、ややこしいので省略）。③RNAME（Reference NAME）。マップされたリファレンスの配列名情報。マップされなかったら*。④POS（start POSision）。マップされたリードの開始点。⑤MAPQ（MAPping Quality）。マッピングのクオリティ。高いほどよい

	A	B	C									
1	@SQ	SN:chromosome	LN:2277981									
2	@SQ	SN:plasmid1	LN:81630									
3	@SQ	SN:plasmid2	LN:40973									
4	@PG	① ID:bwa	② PN:b	③ VN:0.	④ r104	⑤						
5	DRR024501.1	83	chromosome	565592	60	251M	=	565460	−383	CGCCA(	888EEA??8	XT:A:U
6	DRR024501.1	163	chromosome	565460	60	251M	=	565592	383	ATGAAC	?????B?@<	XT:A:U
7	DRR024501.2	83	chromosome	1188239	60	251M	=	1188115	−375	TCACTT	?0*)'*C?:?(	XT:A:U
8	DRR024501.2	163	chromosome	1188115	60	251M	=	1188239	375	TGGACC	??<??B?9E	XT:A:U
9	DRR024501.3	99	chromosome	721208	29	251M	=	721460	367	CAANGA	===#55<=@	XT:A:U
10	DRR024501.3	147	chromosome	721460	29	136S115M	=	721208	−367	ATTACT	(//(((/(((((/(	XT:A:M
11	DRR024501.4	99	chromosome	236084	60	184M	=	236106	197	AGANAC	???#5<<?D	XT:A:U
12	DRR024501.4	147	chromosome	236106	60	175M	=	236084	−197	TTAATT	ECA?:C:*?:	XT:A:U
13	DRR024501.5	77	*	0	0	*	*	0	0	AGTNCC	?,<#55?,<+<5?<<	
14	DRR024501.5	141	*	0	0	*	*	0	0	GATAAA	,55=,<5<−−5@@@	
15	DRR024501.6	89	chromosome	1861136	37	188M	=	1861136	0	TTTACC	FFFFFFFF[	XT:A:U

W17-2：CIGAR

⑥CIGAR。CIGAR stringと呼ばれるもので、マッピング結果の概要を把握する情報。251Mは、251塩基が一致。136S115MのSは、Soft clippingのこと。136塩基がマップされず、115塩基が（マップされて）一致していたということ。オリジナルのリード長は136+115 = 251塩基の長さ。184Mは、184塩基が一致。入力ファイル中のリード長は(FaQCsでアダプタートリムも行っているので)一定ではなく、このリードは184塩基の長さしかなかったのだろう

⑥

	A	B	C	D								
1	@SQ	SN:chromosome	LN:2277981									
2	@SQ	SN:plasmid1	LN:81630									
3	@SQ	SN:plasmid2	LN:40973									
4	@PG	ID:bwa	PN:bwa	VN:0.6.1-r104								
5	DRR024501.1	83	chromosome	565592	60	251M	=	565460	-383	CGCCA(	888EEA??8	XT:A:U
6	DRR024501.1	163	chromosome	565460	60	251M	=	565592	383	ATGAAC	?????B?@<	XT:A:U
7	DRR024501.2	83	chromosome	1188239	60	251M	=	1188115	-375	TCACTT	?0*)'*C?:?(	XT:A:U
8	DRR024501.2	163	chromosome	1188115	60	251M	=	1188239	375	TGGACC	??<??B?9E	XT:A:U
9	DRR024501.3	99	chromosome	721208	29	251M	=	721460	367	CAANGA	===#55<=@	XT:A:U
10	DRR024501.3	147	chromosome	721460	29	136S115M	=	721208	-367	ATTACT	(//(((/((((/(	XT:A:M
11	DRR024501.4	99	chromosome	236084	60	184M	=	236106	197	AGANAC	???#5<<?D	XT:A:U
12	DRR024501.4	147	chromosome	236106	60	175M	=	236084	-197	TTAATT	ECA?:C:*?:	XT:A:U
13	DRR024501.5	77	*	0	0	*	*	0	0	AGTNCC	?,<#55?,<+<5?<<	
14	DRR024501.5	141	*	0	0	*	*	0	0	GATAAA	,55=,<5<--5@@@	
15	DRR024501.6	89	chromosome	1861136	37	188M	=	1861136	0	TTTACC	EEEEEEEE[	XT:A:U

W17-3：RNEXTとPNEXT

⑦RNEXT。*は、マップされなかったもの。=は、paired-endでもう片方が同じリファレンス配列にマップされた場合。もし異なっていれば配列名(chromosome, plasmid1, plasmid2)となる。⑧PNEXTは、paired-endのもう片方のマップされた座標情報

	A	B	C	D	E	F					
1	@SQ	SN:chromosome	LN:2277981								
2	@SQ	SN:plasmid1	LN:81630								
3	@SQ	SN:plasmid2	LN:40973								
4	@PG	ID:bwa	PN:b ③	VN:0.6.1-r104			⑦	⑧			
5	DRR024501.1	83	chromosome	565592	60	251M	=	565460	-383	CGCCA(	888EEA??8 XT:A:U
6	DRR024501.1	163	chromosome	565460	60	251M	=	565592	383	ATGAAC	?????B?@< XT:A:U
7	DRR024501.2	83	chromosome	1188239	60	251M	=	1188115	-375	TCACTT	?0*)'*C?:?(XT:A:U
8	DRR024501.2	163	chromosome	1188115	60	251M	=	1188239	375	TGGACC	??<??B?9E XT:A:U
9	DRR024501.3	99	chromosome	721208	29	251M	=	721460	367	CAANGA	===#55<=@ XT:A:U
10	DRR024501.3	147	chromosome	721460	29	136S115M	=	721208	-367	ATTACT	(//(((/(((((/(XT:A:M
11	DRR024501.4	99	chromosome	236084	60	184M	=	236106	197	AGANAC	???#5<<?D XT:A:U
12	DRR024501.4	147	chromosome	236106	60	175M	=	236084	-197	TTAATT	ECA?:C:*?: XT:A:U
13	DRR024501.5	77	*	0	0	*	*	0	0	AGTNCC	?,<#55?,<+<5?<<
14	DRR024501.5	141	*	0	0	*	*	0	0	GATAAA	,55=,<5<--5@@@
15	DRR024501.6	89	chromosome	1861136	37	188M	=	1861136	0	TTTACC	EEEEEEEE XT:A:U

つまりこんな感じの関係です

W17-3：RNEXTとPNEXT

	A	B	C	D	E	F	G	H	I	J	K	L
1	@SQ	SN:chromosome	LN:2277981									
2	@SQ	SN:plasmid1	LN:81630									
3	@SQ	SN:plasmid2	LN:40973									
4	@PG	ID:bwa	PN:b	VN:0.	r104							
5	DRR024501.1	83	chromosome	565592	60	251M	=	565460	-383	CGCCA(	888EEA??8	XT:A:U
6	DRR024501.1	163	chromosome	565460	60	251M	=	565592	383	ATGAAC	?????B?@<	XT:A:U
7	DRR024501.2	83	chromosome	1188239	60	251M	=	1188115	-375	TCACTT	?0*)'*C?:?(	XT:A:U
8	DRR024501.2	163	chromosome	1188115	60	251M	=	1188239	375	TGGACC	??<??B?9E	XT:A:U
9	DRR024501.3	99	chromosome	721208	29	251M	=	721460	367	CAANGA	===#55<=@	XT:A:U
10	DRR024501.3	147	chromosome	721460	29	136S115M	=	721208	-367	ATTACT	(//(((/((((/(	XT:A:N
11	DRR024501.4	99	chromosome	236084	60	184M	=	236106	197	AGANAC	???#5<<?D	XT:A:U
12	DRR024501.4	147	chromosome	236106	60	175M	=	236084	-197	TTAATT	ECA?:C:*?:	XT:A:U
13	DRR024501.5	77	*	0	0	*	*	0	0	AGTNCC	?,<#55?,<+<5?<<	
14	DRR024501.5	141	*	0	0	*	*	0	0	GATAAA	,55=,<5<--5@@@	
15	DRR024501.6	89	chromosome	1861136	37	188M	=	1861136	0	TTTACC	FFFFFFFF	XT:A:U

W17-3：RNEXTとPNE

⑦のRNEXTが=や*でないものの例。sequence1をqueryとしてBLAST実行した結果にも、plasmid1 (sequence2)とplasmid2 (sequence3)にヒットしたものがあったので妥当。W7-4

③ ④ ⑦ ⑧

	A	B	C	D	E	F	G	H	I	J	K	L
8618	DRR024501.4314	163	chromosome	1352183	60	251M	=	1352317	385	CAACGA	????<?<?B	XT:A:U
8619	DRR024501.4315	99	chromosome	87242	60	251M	=	87383	392	CAACTA	AAAA?BBE	XT:A:U
8620	DRR024501.4315	147	chromosome	87383	60	251M	=	87242	-392	GCAATT	*C:0.'2.*:1(	XT:A:U
8621	DRR024501.4316	83	chromosome	1523909	60	251M	=	1523769	-391	ATCGAC	.EGGECCE	XT:A:U
8622	DRR024501.4316	163	chromosome	1523769	60	251M	=	1523909	391	TCAGGT	?????BBB[	XT:A:U
8623	DRR024501.4317	113	plasmid2	5400	37	251M	ch?	668425	0	ATAAAT	*EE?CEC:(	XT:A:U
8624	DRR024501.4317	177	chromosome	668425	37	251M	plas	5400	0	AATAGT	0:*:'?//*0*	XT:A:U
8625	DRR024501.4318	83	chromosome	1698448	60	251M	=	1698316	-383	GACCG(	>?849CGE	XT:A:U
8626	DRR024501.4318	163	chromosome	1698316	60	251M	=	1698448	383	CTAGTT	?????BB?D	XT:A:U
8627	DRR024501.4319	83	chromosome	1401590	29	251M	=	1401462	-379	ATTGGT	*EGECGE(	XT:A:U
8628	DRR024501.4319	163	chromosome	1401462	29	128M123S	=	1401590	379	CCACG(	????ABB@I	XT:A:M
8629	DRR024501.4320	69	chromosome	1220366	0	*	=	1220366	0	GTATCT	13;?*0:A.8?CCC?	
8630	DRR024501.4320	137	chromosome	1220366	37	109M	=	1220366	0	TGATTA	?????BBB[	XT:A:U
8631	DRR024501.4321	83	chromosome	873625	60	251M	=	873528	-348	GGCTTC	088A8>CE	XT:A:U
8632	DRR024501.4321	163	chromosome	873528	60	251M	=	873625	348	CTAAAT	?????BB?E	XT:A:U
8633	DRR024501.4322	77	*	0	0	*	*	0	0	GTATCT	FBFB5BFF@DEEE	
8634	DRR024501.4322	141	*	0	0	*	*	0	0	TCTCGC	++4:@@99@1*109	

W17-3：RNEXTとPNEXT

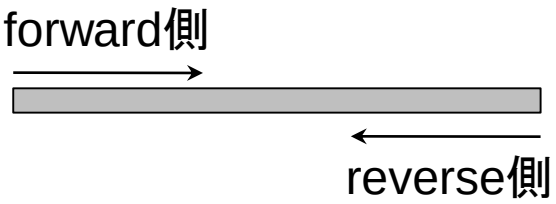
⑦のRNEXTが=の例ですが、なんだかよくわからないものもありますね。これは、、、ペアのリードが同じ場所にマップされたということでしょうかね。。。

	A	B	C	D	E	F	G	H	I	J	K	L
8618	DRR024501.4314	163	chromosome	1352183	60	251M	=	1352317	385	CAACGA	????<?<?B	XT:A:U
8619	DRR024501.4315	99	chromosome	87242	60	251M	=	87383	392	CAACTA	AAAA?BBE	XT:A:U
8620	DRR024501.4315	147	chromosome	87383	60	251M	=	87242	-392	GCAATT	*C:0.'2.*:1(	XT:A:U
8621	DRR024501.4316	83	chromosome	1523909	60	251M	=	1523769	-391	ATCGAC	.EGGECCE	XT:A:U
8622	DRR024501.4316	163	chromosome	1523769	60	251M	=	1523909	391	TCAGGT	?????BBB[	XT:A:U
8623	DRR024501.4317	113	plasmid2	5400	37	251M	chr	668425	0	ATAAAT	*EE?CEC:(	XT:A:U
8624	DRR024501.4317	177	chromosome	668425	37	251M	plas	5400	0	AATAGT	0:*:'?//*0*	XT:A:U
8625	DRR024501.4318	83	chromosome	1698448	60	251M	=	1698316	-383	GACCG(	>?849CGE	XT:A:U
8626	DRR024501.4318	163	chromosome	1698316	60	251M	=	1698448	383	CTAGTT	?????BB?D	XT:A:U
8627	DRR024501.4319	83	chromosome	1401590	29	251M	=	1401462	-379	ATTGGT	*EGECGE(	XT:A:U
8628	DRR024501.4319	163	chromosome	1401462	29	128M123S	=	1401590	379	CCACG(	????ABB@I	XT:A:M
8629	DRR024501.4320	69	chromosome	1220366	0	*	=	1220366	0	GTATCT	13;?*0:A.8?CCC?	
8630	DRR024501.4320	137	chromosome	1220366	37	109M	=	1220366	0	TGATTA	?????BBB[	XT:A:U
8631	DRR024501.4321	83	chromosome	873625	60	251M	=	873528	-348	GGCTTC	088A8>CE	XT:A:U
8632	DRR024501.4321	163	chromosome	873528	60	251M	=	873625	348	CTAAAT	?????BB?E	XT:A:U
8633	DRR024501.4322	77	*	0	0	*	*	0	0	GTATCT	FBFB5BFF@DEEE	
8634	DRR024501.4322	141	*	0	0	*	*	0	0	TCTCGC	++4:@@99@1*109	

W17-4：TLEN

⑨TLEN（observed Template LENgth）。paired-endでシークエンスする前のテンプレート断片配列の長さ。概ね、「(⑧－④)の絶対値 + ⑥」のような感じで計算されます

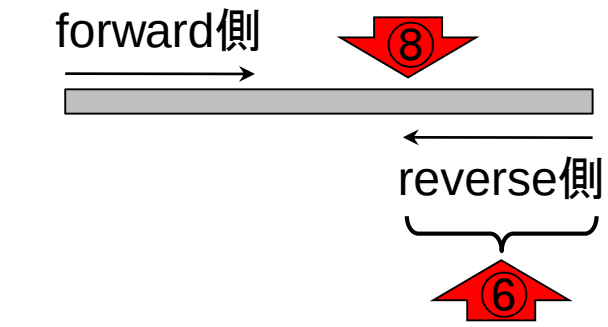
	A	B	C	D	E	F	G	H	I	J	K	L
1	@SQ	SN:chromosome	LN:2277981									
2	@SQ	SN:plasmid1	LN:81630									
3	@SQ	SN:plasmid2	LN:40973									
4	@PG	ID:bwa	PN:bwa	VN:0.[illegible]r104		⑥		⑧	⑨			
5	DRR024501.1	83	chromosome	565592	60	251M	=	565460	−383	CGCCA(	888EEA??8	XT:A:U
6	DRR024501.1	163	chromosome	565460	60	251M	=	565592	383	ATGAAC	?????B?@<	XT:A:U
7	DRR024501.2	83	chromosome	1188239	60	251M	=	1188115	−375	TCACTT	?0*)'*C?:?(	XT:A:U
8	DRR024501.2	163	chromosome	1188115	60	251M	=	1188239	375	TGGACC	??<??B?9E	XT:A:U
9	DRR024501.3	99	chromosome	721208	29	251M	=	721460	367	CAANGA	===#55<=@	XT:A:U
10	DRR024501.3	147	chromosome	721460	29	136S115M	=	721208	−367	ATTACT	(//(((/((((/(	XT:A:M
11	DRR024501.4	99	chromosome	236084	60	184M	=	236106	197	AGANAC	???#5<<?D	XT:A:U
12	DRR024501.4	147	chromosome	236106	60	175M	=	236084	−197	TTAATT	ECA?:C:*?:	XT:A:U
13	DRR024501.5	77	*	0	0	*	*	0	0	AGTNCC	?,<#55?,<+<5?<<	
14	DRR024501.5	141	*	0	0	*	*	0	0	GATAAA	,55=,<5<−−5@@@	
15	DRR024501.6	89	chromosome	1861136	37	188M	=	1861136	0	TTTACC	FFFFFFFF	XT:A:U



W17-4：TLEN

例えば、⑨ = |565460 − 565592| + 251 = 383のように計算します。④と⑧は入れ替わってもいいので、⑨の値がプラスかマイナスかは本質的な問題ではありません。⑥はリードの長さです。アダプタートリムされているものもあるので一定ではありません

	A	B	C	D	E	F	G	H	I	J	K	L
1	@SQ	SN:chromosome	LN:2277981									
2	@SQ	SN:plasmid1	LN:81630									
3	@SQ	SN:plasmid2	LN:40973									
4	@PG	ID:bwa	PN:bwa	VN:0.[illegible]r104								
5	DRR024501.1	83	chromosome	565592	60	251M	=	565460	−383	CGCCA(	888EEA??8	XT:A:U
6	DRR024501.1	163	chromosome	565460	60	251M	=	565592	383	ATGAAC	?????B?@<	XT:A:U
7	DRR024501.2	83	chromosome	1188239	60	251M	=	1188115	−375	TCACTT	?0*)'*C?:?(	XT:A:U
8	DRR024501.2	163	chromosome	1188115	60	251M	=	1188239	375	TGGACC	??<??B?9E	XT:A:U
9	DRR024501.3	99	chromosome	721208	29	251M	=	721460	367	CAANGA	===#55<=@	XT:A:U
10	DRR024501.3	147	chromosome	721460	29	136S115M	=	721208	−367	ATTACT	(//(((/((((/(	XT:A:M
11	DRR024501.4	99	chromosome	236084	60	184M	=	236106	197	AGANAC	???#5<<?D	XT:A:U
12	DRR024501.4	147	chromosome	236106	60	175M	=	236084	−197	TTAATT	ECA?:C:*?:	XT:A:U
13	DRR024501.5	77	*	0	0	*	*	0	0	AGTNCC	?,<#55?,<+<5?<<	
14	DRR024501.5	141	*	0	0	*	*	0	0	GATAAA	,55=,<5<−−5@@@	
15	DRR024501.6	89	chromosome	1861136	37	[illegible]	=	1861136	0	TTTACC	EEEEEEEE	XT:A:U



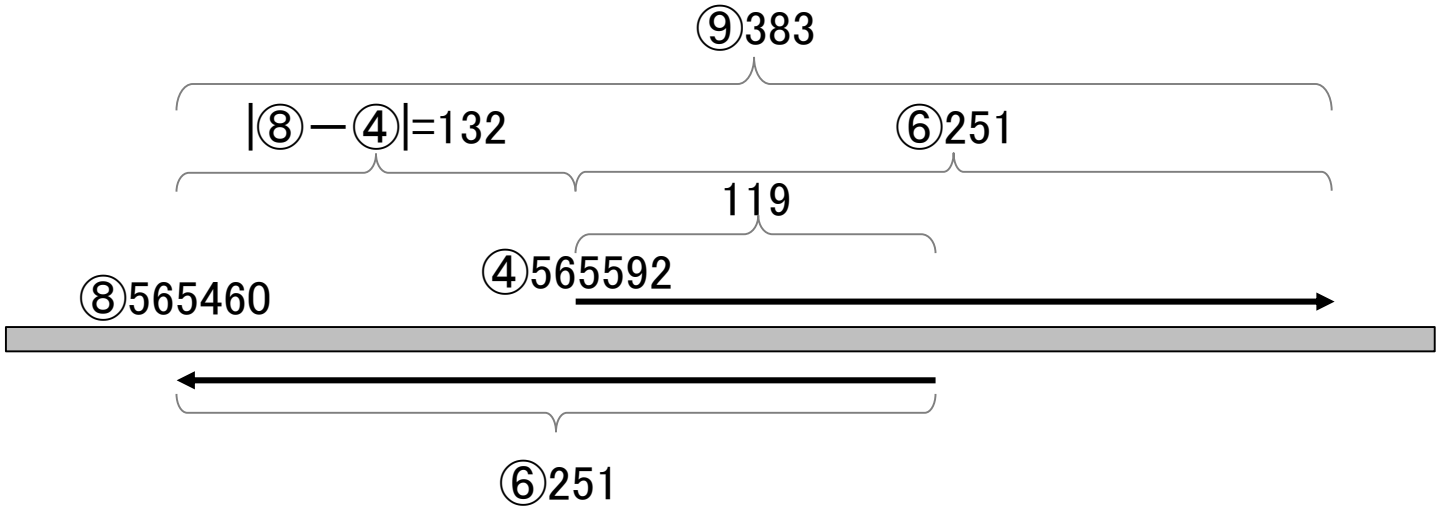
もう少し丁寧に示すとこんな感じです

W17-4：TLEN

④ POS (start **POS**ision)
⑥ CIGAR (CIGAR string)
⑧ PNEXT (paired-endのもう片方のマップされた座標情報)
⑨ TLEN (observed **T**emplate **LEN**gth)＝｜⑧－④｜＋⑥

	A	B	C	D	E	F	G	H	I	J	K	L
1	@SQ	SN:chromosome	LN:2277981									
2	@SQ	SN:plasmid1	LN:81630									
3	@SQ	SN:plasmid2	LN:40973									
4	@PG	ID:bwa	PN:bwa	VN:0.[illegible]r104								
5	DRR024501.1	83	chromosome	565592	60	251M	=	565460	-383	CGCCA[illegible]	888EEA??8	XT:A:[illegible]
6	DRR024501.1	163	chromosome	565460	60	251M	=	565592	383	ATGAAC	?????B?@<	XT:A:[illegible]

④ ⑥ ⑧ ⑨



W17-5：SEQとQUAL

⑩SEQ（SEQuence）。リード塩基配列情報。⑪QUAL（QUALity）。1文字表記のクオリティスコア

	A	B	C	D	E	F	G	H	I	J	K	L
1	@SQ	SN:chromosome	LN:2277981									
2	@SQ	SN:plasmid1	LN:81630									
3	@SQ	SN:plasmid2	LN:40973									
4	@PG	ID:bwa	PN:bwa	VN:0.6.1−r104						⑩	⑪	
5	DRR024501.1	83	chromosome	565592	60	251M	=	565460	−383	CGCCA(	888EEA??8	XT:A:U
6	DRR024501.1	163	chromosome	565460	60	251M	=	565592	383	ATGAAC	?????B?@<	XT:A:U
7	DRR024501.2	83	chromosome	1188239	60	251M	=	1188115	−375	TCACTT	?0*)'*C?:?(	XT:A:U
8	DRR024501.2	163	chromosome	1188115	60	251M	=	1188239	375	TGGACC	??<??B?9E	XT:A:U
9	DRR024501.3	99	chromosome	721208	29	251M	=	721460	367	CAANGA	===#55<=@	XT:A:U
10	DRR024501.3	147	chromosome	721460	29	136S115M	=	721208	−367	ATTACT	(//(((/((((/(	XT:A:M
11	DRR024501.4	99	chromosome	236084	60	184M	=	236106	197	AGANAC	???#5<<?D	XT:A:U
12	DRR024501.4	147	chromosome	236106	60	175M	=	236084	−197	TTAATT	ECA?:C:*?:	XT:A:U
13	DRR024501.5	77	*	0	0	*	*	0	0	AGTNCC	?,<#55?,<+<5?<<	
14	DRR024501.5	141	*	0	0	*	*	0	0	GATAAA	,55=,<5<−−5@@@	
15	DRR024501.6	89	chromosome	1861136	37	188M	=	1861136	0	TTTACC	EEEEEEEE	XT:A:U

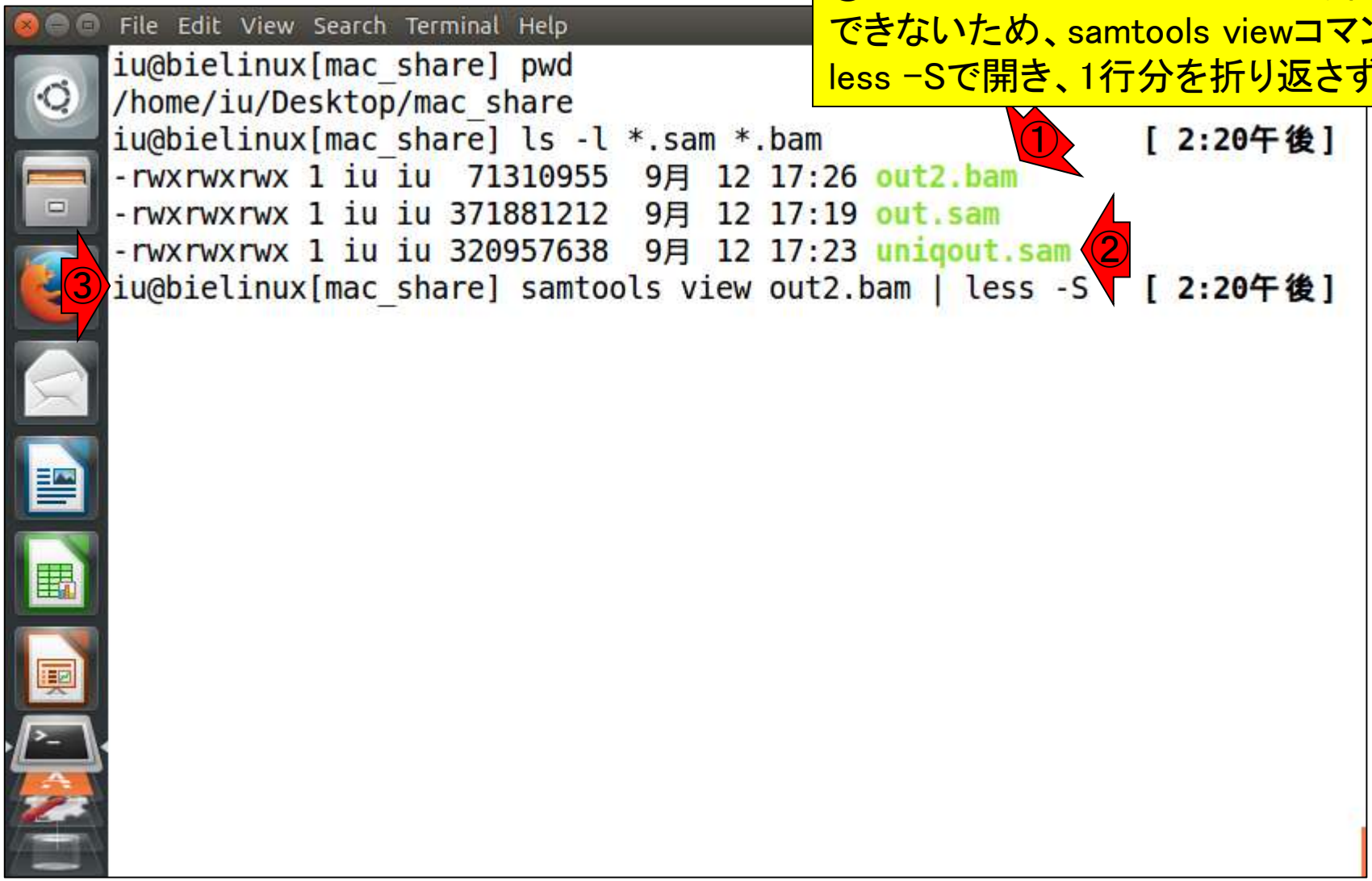
W17-5：optional field

⑫12列目以降は付加的な情報を含む場所（optional field）。ここが「XT:A:U」となっているものがユニークヒットに相当するもの。uniqout.samは、この条件を満たすリードペアのみからなる

	A	B	C	D	E	F	G	H	I	J	K	L
1	@SQ	SN:chromosome	LN:2277981									
2	@SQ	SN:plasmid1	LN:81630									
3	@SQ	SN:plasmid2	LN:40973									
4	@PG	ID:bwa	PN:bwa	VN:0.6.1−r104								⑫
5	DRR024501.1	83	chromosome	565592	60	251M	=	565460	−383	CGCCA(	888EEA??8	XT:A:U
6	DRR024501.1	163	chromosome	565460	60	251M	=	565592	383	ATGAAC	?????B?@<	XT:A:U
7	DRR024501.2	83	chromosome	1188239	60	251M	=	1188115	−375	TCACTT	?0*)'*C?:?(	XT:A:U
8	DRR024501.2	163	chromosome	1188115	60	251M	=	1188239	375	TGGACC	??<??B?9E	XT:A:U
9	DRR024501.3	99	chromosome	721208	29	251M	=	721460	367	CAANGA	===#55<=@	XT:A:U
10	DRR024501.3	147	chromosome	721460	29	136S115M	=	721208	−367	ATTACT	(//(((/((((/(	XT:A:N
11	DRR024501.4	99	chromosome	236084	60	184M	=	236106	197	AGANAC	???#5<<?D	XT:A:U
12	DRR024501.4	147	chromosome	236106	60	175M	=	236084	−197	TTAATT	ECA?:C:*?:	XT:A:U
13	DRR024501.5	77	*	0	0	*	*	0	0	AGTNCC	?,<#55?,<+<5?<<	
14	DRR024501.5	141	*	0	0	*	*	0	0	GATAAA	,55=,<5<−−5@@@	
15	DRR024501.6	89	chromosome	1861136	37	188M	=	1861136	0	TTTACC	FFFFFFFF[	XT:A:U

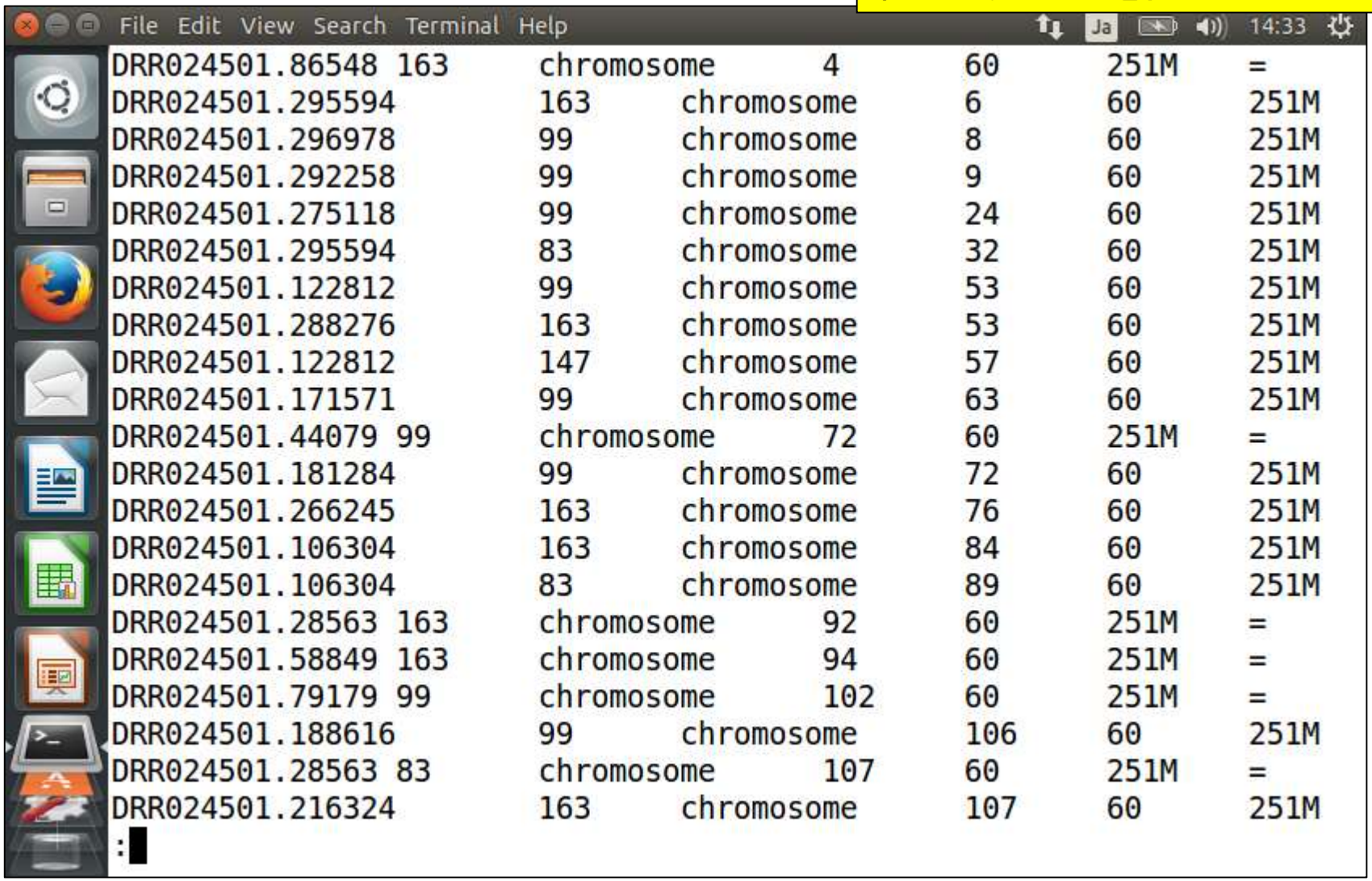
W17-6：out2.bam

①out2.bamは、out.samからユニーク化処理後のSAMファイルである②uniqout.samをベースに作成。③out2.bamはバイナリファイル。直接眺めることはできないため、samtools viewコマンドで見ている。less -Sで開き、1行分を折り返さずに表示（W16-6）



W17-6：out2.bam

こんな感じ。若干ずれていますが、chromosomeの右隣の数値が昇順にソートされていることがわかる。Gキーを押してページの最終行に移動



W17-6：out2.bam

out2.bamの最終行はこんな感じ。plasmid2は40,973塩基なので(W16-6)、①最後のほうが40786で終わっているのは妥当といえる。尚、out2.bam作成過程でマップされなかったリードは除外されているので、このファイル中には存在しない

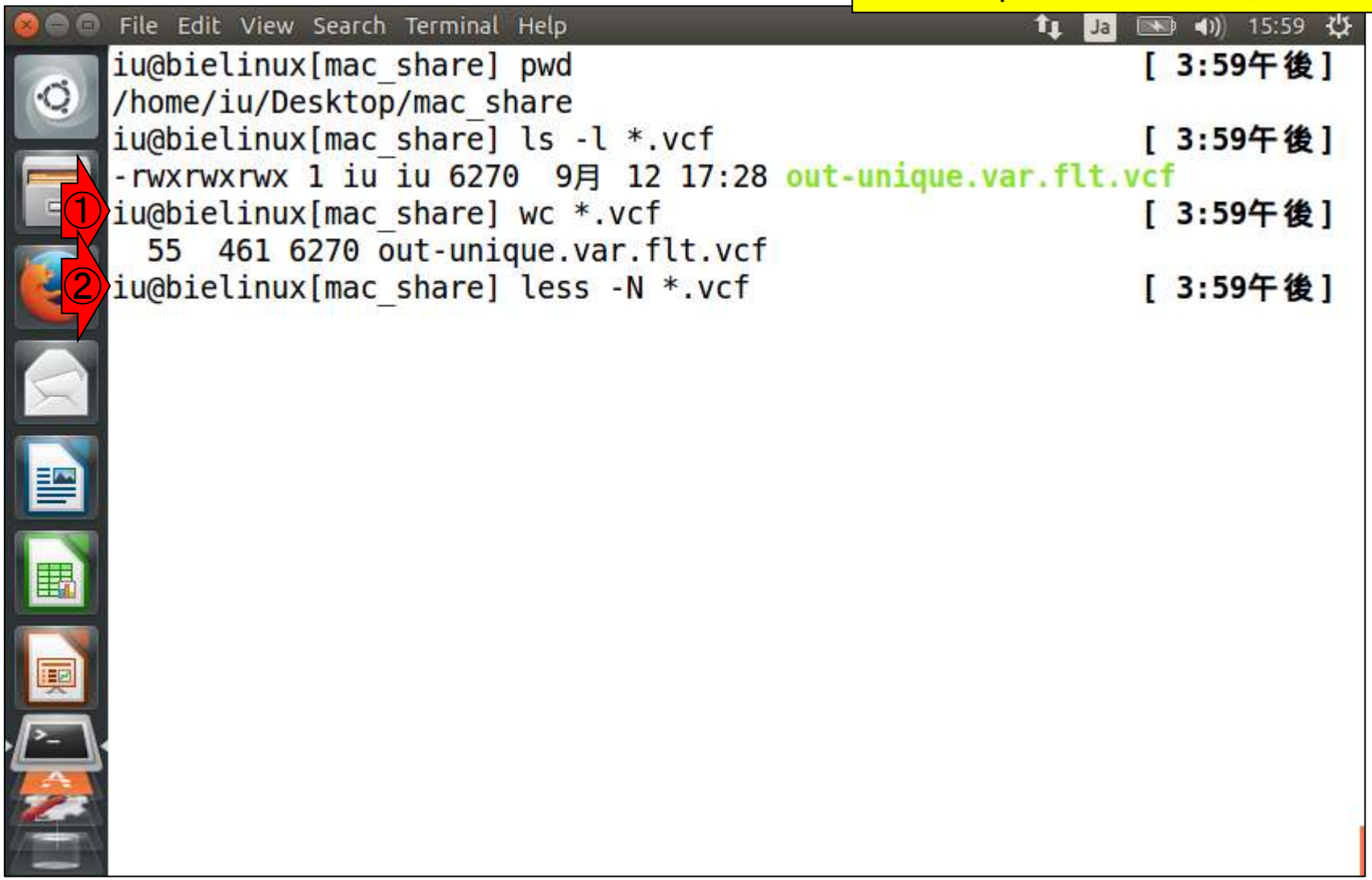
File Edit View Search Terminal Help

```
DRR024501.228401        83      plasmid2                40680   60      251M
DRR024501.178522        147     plasmid2                40680   60      251M
DRR024501.240655        83      plasmid2                40680   60      251M
DRR024501.73398 147     plasmid2        40685   60      251M    =
DRR024501.267503        147     plasmid2                40688   60      251M
DRR024501.254114        145     plasmid2                40694   37      137M
DRR024501.254114        97      plasmid2                40695   37      132M
DRR024501.264041        147     plasmid2                40695   60      251M
DRR024501.147571        83      plasmid2                40706   60      251M
DRR024501.141761        147     plasmid2                40707   60      251M
DRR024501.218771        83      plasmid2                40708   60      251M
DRR024501.217892        147     plasmid2                40710   60      251M
DRR024501.165704        83      plasmid2                40714   60      251M
DRR024501.13639 83      plasmid2        40717   60      251M    =
DRR024501.127223        147     plasmid2                40718   60      251M
DRR024501.176446        99      plasmid2                40723   60      130M
DRR024501.176446        147     plasmid2                40741   60      113M
DRR024501.103552        99      plasmid2                40772   60      130M
DRR024501.103552        147     plasmid2                40781   60      128M
DRR024501.271587        145     plasmid2                40785   37      138M
DRR024501.271587        97      plasmid2                40786   37      123M
(END)
```

①

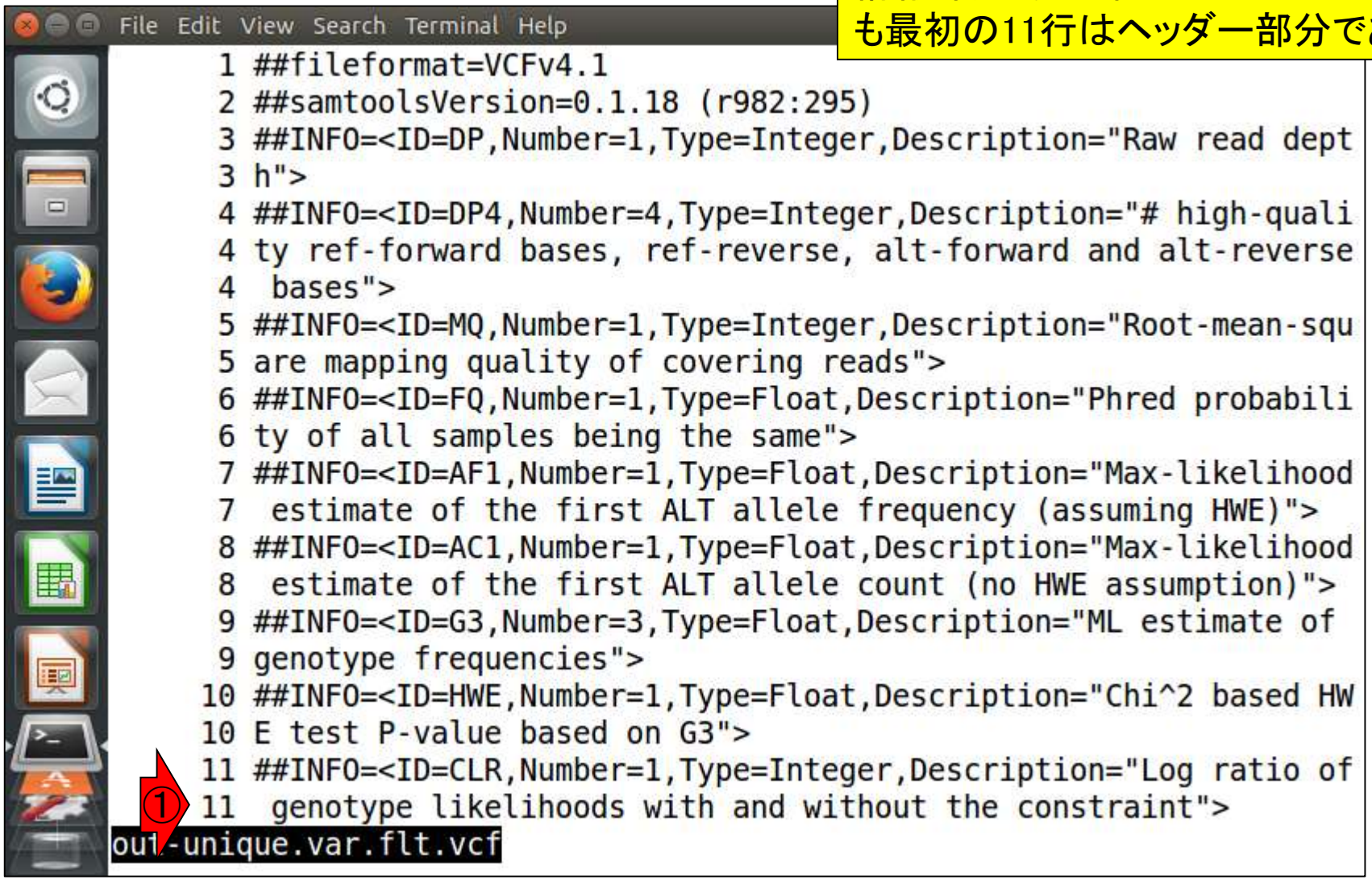
W18-1：VCFファイル

VCF (Variant Call Format)形式ファイルの説明。
DDBJ Pipelineからダウンロードして解凍した、①out-unique.var.flt.vcfは、55行。②less -Nで確認



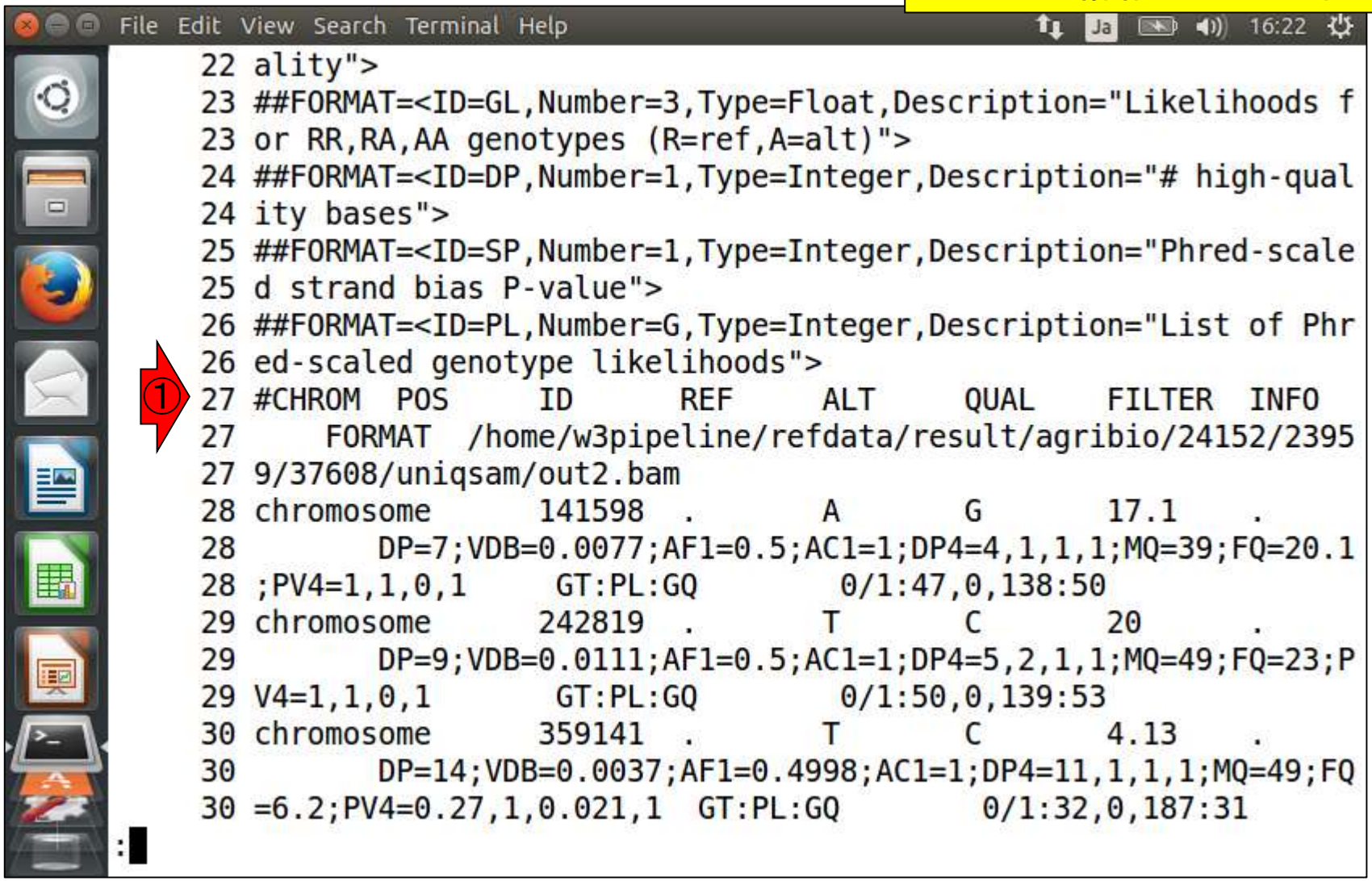
W18-1：VCFファイル

less -Nで確認しているところ。ここで見えているのは、行頭が#から始まるヘッダー行（メタデータ情報部分）。全部で55行しかないのに、①少なくとも最初の11行はヘッダー部分であることがわかる



W18-1：VCFファイル

下矢印キーを押してページ下部に移動。①27行目がヘッダー行の最後。ここの情報が28行目以降の情報の列名に相当するのだろう



W18-2：Excel

Excelで眺めるとこんな感じ。1行目はVCFのバージョン。これはver. 4.1のようだ

	A	B	C	D	E	F	G	H	I	J	K	L
1	##fileformat=VCFv4.1											
2	##samtoolsVersion=0.1.18 (r982:295)											
3	##INFO=<ID=DP,Number=1,Type=Integer,Description="Raw read depth">											
4	##INFO=<ID=DP4,Number=4,Type=Integer,Description="# high-quality ref-forward bases, ref-reverse, alt											
5	##INFO=<ID=MQ,Number=1,Type=Integer,Description="Root-mean-square mapping quality of covering re											
6	##INFO=<ID=FQ,Number=1,Type=Float,Description="Phred probability of all samples being the same">											
7	##INFO=<ID=AF1,Number=1,Type=Float,Description="Max-likelihood estimate of the first ALT allele frequ											
8	##INFO=<ID=AC1,Number=1,Type=Float,Description="Max-likelihood estimate of the first ALT allele coun											
9	##INFO=<ID=G3,Number=3,Type=Float,Description="ML estimate of genotype frequencies">											
10	##INFO=<ID=HWE,Number=1,Type=Float,Description="Chi^2 based HWE test P-value based on G3">											
11	##INFO=<ID=CLR,Number=1,Type=Integer,Description="Log ratio of genotype likelihoods with and withou											
12	##INFO=<ID=UGT,Number=1,Type=String,Description="The most probable unconstrained genotype config											
13	##INFO=<ID=CGT,Number=1,Type=String,Description="The most probable constrained genotype configura											
14	##INFO=<ID=PV4,Number=4,Type=Float,Description="P-values for strand bias, baseQ bias, mapQ bias and											

W18-2：Excel

27行目以降の主要な結果の全体像。①REFは、概要配列側の塩基。②ALTは、変異塩基(ALTanative base)。③QUALは、クオリティスコア（高いほどよい）

	A	B	C	①	②	③	G	H	I	J
27	#CHROM	POS	ID	REF	ALT	QUAL	FILTER	INFO	FORMAT	/home/w3pipeline/refdata/
28	chromosome	141598	.	A	G	17.1	.	DP=7;VDB=0.0077;AF1=0.5;AC1	GT:PL:GQ	0/1:47,0,138:50
29	chromosome	242819	.	T	C	20	.	DP=9;VDB=0.0111;AF1=0.5;AC1	GT:PL:GQ	0/1:50,0,139:53
30	chromosome	359141	.	T	C	4.13	.	DP=14;VDB=0.0037;AF1=0.4998	GT:PL:GQ	0/1:32,0,187:31
31	chromosome	463882	.	T	G	6.2	.	DP=13;VDB=0.0074;AF1=0.4999	GT:PL:GQ	0/1:35,0,216:36
32	chromosome	660031	.	C	T	8.64	.	DP=11;VDB=0.0063;AF1=0.5;AC	GT:PL:GQ	0/1:38,0,197:40
33	chromosome	792963	.	T	C	4.13	.	DP=14;VDB=0.0153;AF1=0.4998	GT:PL:GQ	0/1:32,0,200:31
34	chromosome	887422	.	G	A	10.4	.	DP=11;VDB=0.0070;AF1=0.5;AC	GT:PL:GQ	0/1:40,0,188:42
35	chromosome	913470	.	T	C	14.2	.	DP=9;VDB=0.0056;AF1=0.5;AC1	GT:PL:GQ	0/1:44,0,144:47
36	chromosome	965171	.	T	C	17.1	.	DP=20;VDB=0.0071;AF1=0.5;AC	GT:PL:GQ	0/1:47,0,249:50
37	chromosome	1134950	.	A	G	7.8	.	DP=23;VDB=0.0076;AF1=0.5;AC	GT:PL:GQ	0/1:37,0,255:39
38	chromosome	1160048	.	A	G	29	.	DP=17;VDB=0.0054;AF1=0.5;AC	GT:PL:GQ	0/1:59,0,255:62
39	chromosome	1160095	.	A	G	19.1	.	DP=10;VDB=0.0033;AF1=0.5;AC	GT:PL:GQ	0/1:49,0,191:52
40	chromosome	1240248	.	GCCCC	GCCCCC	214	.	INDEL;DP=48;VDB=0.0137;AF1=	GT:PL:GQ	1/1:255,128,0:99
41	chromosome	1455552	.	CAAAA	CAAAAA	214	.	INDEL;DP=58;VDB=0.0164;AF1=	GT:PL:GQ	1/1:255,130,0:99
42	chromosome	1475721	.	T	C	52	.	DP=19;VDB=0.0021;AF1=0.5;AC	GT:PL:GQ	0/1:82,0,246:85
43	chromosome	1752680	.	G	A	29	.	DP=24;VDB=0.0129;AF1=0.5;AC	GT:PL:GQ	0/1:59,0,255:62
44	chromosome	1941380	.	T	C	7.8	.	DP=13;VDB=0.0046;AF1=0.5;AC	GT:PL:GQ	0/1:37,0,214:39
45	chromosome	2052955	.	A	G	12.3	.	DP=8;VDB=0.0084;AF1=0.5;AC1	GT:PL:GQ	0/1:42,0,103:44
46	chromosome	2054084	.	TC	T	13.7	.	INDEL;DP=8;VDB=0.0009;AF1=0	GT:PL:GQ	0/1:51,0,174:54
47	chromosome	2056279	.	A	G	24	.	DP=6;VDB=0.0057;AF1=0.5;AC1	GT:PL:GQ	0/1:54,0,113:57
48	chromosome	2057564	.	T	C	8.64	.	DP=10;VDB=0.0041;AF1=0.5;AC	GT:PL:GQ	0/1:38,0,129:40
49	chromosome	2058145	.	C	T,G	108	.	DP=14;VDB=0.0147;AF1=1;AC1=	GT:PL:GQ	1/1:141,39,0,141,28,138:75
50	plasmid1	37177	.	G	T	54	.	DP=49;VDB=0.0087;AF1=0.5;AC	GT:PL:GQ	0/1:84,0,255:87
51	plasmid1	66609	.	A	G	225	.	DP=55;VDB=0.0060;AF1=0.5;AC	GT:PL:GQ	0/1:255,0,255:99
52	plasmid1	66637	.	C	T	225	.	DP=71;VDB=0.0141;AF1=0.5;AC	GT:PL:GQ	0/1:255,0,249:99
53	plasmid1	66668	.	C	T	225	.	DP=80;VDB=0.0153;AF1=0.5;AC	GT:PL:GQ	0/1:255,0,255:99
54	plasmid2	2569	.	CAAAAA	CAAAA	177	.	INDEL;DP=110;VDB=0.0166;AF1	GT:PL:GQ	1/1:218,255,0:99
55	plasmid2	37805	.	TAAAAA	TAAAA	104	.	INDEL;DP=31;VDB=0.0058;AF1=	GT:PL:GQ	1/1:145,93,0:99

W18-2：Excel

④INFOは、変異情報。⑤FORMATは、genotype（GT)のフォーマット、⑥は⑤の形式でgenotypeの具体的な情報が記載されている

	A	B	C	D	E	F	G	④	⑤	⑥
27	#CHROM	POS	ID	REF	ALT	QUAL	FILTER	INFO	FORMAT	/home/w3pipeline/refdata/
28	chromosome	141598	.	A	G	17.1	.	DP=7;VDB=0.0077;AF1=0.5;AC1	GT:PL:GQ	0/1:47,0,138:50
29	chromosome	242819	.	T	C	20	.	DP=9;VDB=0.0111;AF1=0.5;AC1	GT:PL:GQ	0/1:50,0,139:53
30	chromosome	359141	.	T	C	4.13	.	DP=14;VDB=0.0037;AF1=0.4998	GT:PL:GQ	0/1:32,0,187:31
31	chromosome	463882	.	T	G	6.2	.	DP=13;VDB=0.0074;AF1=0.4999	GT:PL:GQ	0/1:35,0,216:36
32	chromosome	660031	.	C	T	8.64	.	DP=11;VDB=0.0063;AF1=0.5;AC	GT:PL:GQ	0/1:38,0,197:40
33	chromosome	792963	.	T	C	4.13	.	DP=14;VDB=0.0153;AF1=0.4998	GT:PL:GQ	0/1:32,0,200:31
34	chromosome	887422	.	G	A	10.4	.	DP=11;VDB=0.0070;AF1=0.5;AC	GT:PL:GQ	0/1:40,0,188:42
35	chromosome	913470	.	T	C	14.2	.	DP=9;VDB=0.0056;AF1=0.5;AC1	GT:PL:GQ	0/1:44,0,144:47
36	chromosome	965171	.	T	C	17.1	.	DP=20;VDB=0.0071;AF1=0.5;AC	GT:PL:GQ	0/1:47,0,249:50
37	chromosome	1134950	.	A	G	7.8	.	DP=23;VDB=0.0076;AF1=0.5;AC	GT:PL:GQ	0/1:37,0,255:39
38	chromosome	1160048	.	A	G	29	.	DP=17;VDB=0.0054;AF1=0.5;AC	GT:PL:GQ	0/1:59,0,255:62
39	chromosome	1160095	.	A	G	19.1	.	DP=10;VDB=0.0033;AF1=0.5;AC	GT:PL:GQ	0/1:49,0,191:52
40	chromosome	1240248	.	GCCCC	GCCCCC	214	.	INDEL;DP=48;VDB=0.0137;AF1=	GT:PL:GQ	1/1:255,128,0:99
41	chromosome	1455552	.	CAAAA	CAAAAA	214	.	INDEL;DP=58;VDB=0.0164;AF1=	GT:PL:GQ	1/1:255,130,0:99
42	chromosome	1475721	.	T	C	52	.	DP=19;VDB=0.0021;AF1=0.5;AC	GT:PL:GQ	0/1:82,0,246:85
43	chromosome	1752680	.	G	A	29	.	DP=24;VDB=0.0129;AF1=0.5;AC	GT:PL:GQ	0/1:59,0,255:62
44	chromosome	1941380	.	T	C	7.8	.	DP=13;VDB=0.0046;AF1=0.5;AC	GT:PL:GQ	0/1:37,0,214:39
45	chromosome	2052955	.	A	G	12.3	.	DP=8;VDB=0.0084;AF1=0.5;AC1	GT:PL:GQ	0/1:42,0,103:44
46	chromosome	2054084	.	TC	T	13.7	.	INDEL;DP=8;VDB=0.0009;AF1=0	GT:PL:GQ	0/1:51,0,174:54
47	chromosome	2056279	.	A	G	24	.	DP=6;VDB=0.0057;AF1=0.5;AC1	GT:PL:GQ	0/1:54,0,113:57
48	chromosome	2057564	.	T	C	8.64	.	DP=10;VDB=0.0041;AF1=0.5;AC	GT:PL:GQ	0/1:38,0,129:40
49	chromosome	2058145	.	C	T,G	108	.	DP=14;VDB=0.0147;AF1=1;AC1=	GT:PL:GQ	1/1:141,39,0,141,28,138:75
50	plasmid1	37177	.	G	T	54	.	DP=49;VDB=0.0087;AF1=0.5;AC	GT:PL:GQ	0/1:84,0,255:87
51	plasmid1	66609	.	A	G	225	.	DP=55;VDB=0.0060;AF1=0.5;AC	GT:PL:GQ	0/1:255,0,255:99
52	plasmid1	66637	.	C	T	225	.	DP=71;VDB=0.0141;AF1=0.5;AC	GT:PL:GQ	0/1:255,0,249:99
53	plasmid1	66668	.	C	T	225	.	DP=80;VDB=0.0153;AF1=0.5;AC	GT:PL:GQ	0/1:255,0,255:99
54	plasmid2	2569	.	CAAAAA	CAAAA	177	.	INDEL;DP=110;VDB=0.0166;AF1	GT:PL:GQ	1/1:218,255,0:99
55	plasmid2	37805	.	TAAAAA	TAAAA	104	.	INDEL;DP=31;VDB=0.0058;AF1=	GT:PL:GQ	1/1:145,93,0:99

W18-2：Excel

今回はpaired-endの1サンプルのみを入力としたので、⑥の部分が1列分のみのgenotype情報からなる

	A	B	C	D	E	F	G	H	I	⑥
27	#CHROM	POS	ID	REF	ALT	QUAL	FILTER	INFO	FORMAT	/home/w3pipeline/refdata/
28	chromosome	141598	.	A	G	17.1	.	DP=7;VDB=0.0077;AF1=0.5;AC1	GT:PL:GQ	0/1:47,0,138:50
29	chromosome	242819	.	T	C	20	.	DP=9;VDB=0.0111;AF1=0.5;AC1	GT:PL:GQ	0/1:50,0,139:53
30	chromosome	359141	.	T	C	4.13	.	DP=14;VDB=0.0037;AF1=0.4998	GT:PL:GQ	0/1:32,0,187:31
31	chromosome	463882	.	T	G	6.2	.	DP=13;VDB=0.0074;AF1=0.4999	GT:PL:GQ	0/1:35,0,216:36
32	chromosome	660031	.	C	T	8.64	.	DP=11;VDB=0.0063;AF1=0.5;AC	GT:PL:GQ	0/1:38,0,197:40
33	chromosome	792963	.	T	C	4.13	.	DP=14;VDB=0.0153;AF1=0.4998	GT:PL:GQ	0/1:32,0,200:31
34	chromosome	887422	.	G	A	10.4	.	DP=11;VDB=0.0070;AF1=0.5;AC	GT:PL:GQ	0/1:40,0,188:42
35	chromosome	913470	.	T	C	14.2	.	DP=9;VDB=0.0056;AF1=0.5;AC1	GT:PL:GQ	0/1:44,0,144:47
36	chromosome	965171	.	T	C	17.1	.	DP=20;VDB=0.0071;AF1=0.5;AC	GT:PL:GQ	0/1:47,0,249:50
37	chromosome	1134950	.	A	G	7.8	.	DP=23;VDB=0.0076;AF1=0.5;AC	GT:PL:GQ	0/1:37,0,255:39
38	chromosome	1160048	.	A	G	29	.	DP=17;VDB=0.0054;AF1=0.5;AC	GT:PL:GQ	0/1:59,0,255:62
39	chromosome	1160095	.	A	G	19.1	.	DP=10;VDB=0.0033;AF1=0.5;AC	GT:PL:GQ	0/1:49,0,191:52
40	chromosome	1240248	.	GCCCC	GCCCCC	214	.	INDEL;DP=48;VDB=0.0137;AF1=	GT:PL:GQ	1/1:255,128,0:99
41	chromosome	1455552	.	CAAAA	CAAAAA	214	.	INDEL;DP=58;VDB=0.0164;AF1=	GT:PL:GQ	1/1:255,130,0:99
42	chromosome	1475721	.	T	C	52	.	DP=19;VDB=0.0021;AF1=0.5;AC	GT:PL:GQ	0/1:82,0,246:85
43	chromosome	1752680	.	G	A	29	.	DP=24;VDB=0.0129;AF1=0.5;AC	GT:PL:GQ	0/1:59,0,255:62
44	chromosome	1941380	.	T	C	7.8	.	DP=13;VDB=0.0046;AF1=0.5;AC	GT:PL:GQ	0/1:37,0,214:39
45	chromosome	2052955	.	A	G	12.3	.	DP=8;VDB=0.0084;AF1=0.5;AC1	GT:PL:GQ	0/1:42,0,103:44
46	chromosome	2054084	.	TC	T	13.7	.	INDEL;DP=8;VDB=0.0009;AF1=0	GT:PL:GQ	0/1:51,0,174:54
47	chromosome	2056279	.	A	G	24	.	DP=6;VDB=0.0057;AF1=0.5;AC1	GT:PL:GQ	0/1:54,0,113:57
48	chromosome	2057564	.	T	C	8.64	.	DP=10;VDB=0.0041;AF1=0.5;AC	GT:PL:GQ	0/1:38,0,129:40
49	chromosome	2058145	.	C	T,G	108	.	DP=14;VDB=0.0147;AF1=1;AC1=	GT:PL:GQ	1/1:141,39,0,141,28,138:75
50	plasmid1	37177	.	G	T	54	.	DP=49;VDB=0.0087;AF1=0.5;AC	GT:PL:GQ	0/1:84,0,255:87
51	plasmid1	66609	.	A	G	225	.	DP=55;VDB=0.0060;AF1=0.5;AC	GT:PL:GQ	0/1:255,0,255:99
52	plasmid1	66637	.	C	T	225	.	DP=71;VDB=0.0141;AF1=0.5;AC	GT:PL:GQ	0/1:255,0,249:99
53	plasmid1	66668	.	C	T	225	.	DP=80;VDB=0.0153;AF1=0.5;AC	GT:PL:GQ	0/1:255,0,255:99
54	plasmid2	2569	.	CAAAAA	CAAAA	177	.	INDEL;DP=110;VDB=0.0166;AF1	GT:PL:GQ	1/1:218,255,0:99
55	plasmid2	37805	.	TAAAAA	TAAAA	104	.	INDEL;DP=31;VDB=0.0058;AF1=	GT:PL:GQ	1/1:145,93,0:99

W18-2：Excel

もしBWAマッピング時の入力が3サンプル分だったら、⑥の部分が3列分の情報になる。具体的には、Excelでいうところの K列とL列に2サンプル目と3サンプル目のgenotype情報が追加される

⑥

	A	B	C	D	E	F	G	H	I	/home/w3pipeline/refdata/
27	#CHROM	POS	ID	REF	ALT	QUAL	FILTER	INFO	FORMAT	
28	chromosome	141598	.	A	G	17.1	.	DP=7;VDB=0.0077;AF1=0.5;AC1	GT:PL:GQ	0/1:47,0,138:50
29	chromosome	242819	.	T	C	20	.	DP=9;VDB=0.0111;AF1=0.5;AC1	GT:PL:GQ	0/1:50,0,139:53
30	chromosome	359141	.	T	C	4.13	.	DP=14;VDB=0.0037;AF1=0.4998	GT:PL:GQ	0/1:32,0,187:31
31	chromosome	463882	.	T	G	6.2	.	DP=13;VDB=0.0074;AF1=0.4999	GT:PL:GQ	0/1:35,0,216:36
32	chromosome	660031	.	C	T	8.64	.	DP=11;VDB=0.0063;AF1=0.5;AC	GT:PL:GQ	0/1:38,0,197:40
33	chromosome	792963	.	T	C	4.13	.	DP=14;VDB=0.0153;AF1=0.4998	GT:PL:GQ	0/1:32,0,200:31
34	chromosome	887422	.	G	A	10.4	.	DP=11;VDB=0.0070;AF1=0.5;AC	GT:PL:GQ	0/1:40,0,188:42
35	chromosome	913470	.	T	C	14.2	.	DP=9;VDB=0.0056;AF1=0.5;AC1	GT:PL:GQ	0/1:44,0,144:47
36	chromosome	965171	.	T	C	17.1	.	DP=20;VDB=0.0071;AF1=0.5;AC	GT:PL:GQ	0/1:47,0,249:50
37	chromosome	1134950	.	A	G	7.8	.	DP=23;VDB=0.0076;AF1=0.5;AC	GT:PL:GQ	0/1:37,0,255:39
38	chromosome	1160048	.	A	G	29	.	DP=17;VDB=0.0054;AF1=0.5;AC	GT:PL:GQ	0/1:59,0,255:62
39	chromosome	1160095	.	A	G	19.1	.	DP=10;VDB=0.0033;AF1=0.5;AC	GT:PL:GQ	0/1:49,0,191:52
40	chromosome	1240248	.	GCCCC	GCCCCC	214	.	INDEL;DP=48;VDB=0.0137;AF1=	GT:PL:GQ	1/1:255,128,0:99
41	chromosome	1455552	.	CAAAA	CAAAAA	214	.	INDEL;DP=58;VDB=0.0164;AF1=	GT:PL:GQ	1/1:255,130,0:99
42	chromosome	1475721	.	T	C	52	.	DP=19;VDB=0.0021;AF1=0.5;AC	GT:PL:GQ	0/1:82,0,246:85
43	chromosome	1752680	.	G	A	29	.	DP=24;VDB=0.0129;AF1=0.5;AC	GT:PL:GQ	0/1:59,0,255:62
44	chromosome	1941380	.	T	C	7.8	.	DP=13;VDB=0.0046;AF1=0.5;AC	GT:PL:GQ	0/1:37,0,214:39
45	chromosome	2052955	.	A	G	12.3	.	DP=8;VDB=0.0084;AF1=0.5;AC1	GT:PL:GQ	0/1:42,0,103:44
46	chromosome	2054084	.	TC	T	13.7	.	INDEL;DP=8;VDB=0.0009;AF1=0	GT:PL:GQ	0/1:51,0,174:54
47	chromosome	2056279	.	A	G	24	.	DP=6;VDB=0.0057;AF1=0.5;AC1	GT:PL:GQ	0/1:54,0,113:57
48	chromosome	2057564	.	T	C	8.64	.	DP=10;VDB=0.0041;AF1=0.5;AC	GT:PL:GQ	0/1:38,0,129:40
49	chromosome	2058145	.	C	T,G	108	.	DP=14;VDB=0.0147;AF1=1;AC1=	GT:PL:GQ	1/1:141,39,0,141,28,138:75
50	plasmid1	37177	.	G	T	54	.	DP=49;VDB=0.0087;AF1=0.5;AC	GT:PL:GQ	0/1:84,0,255:87
51	plasmid1	66609	.	A	G	225	.	DP=55;VDB=0.0060;AF1=0.5;AC	GT:PL:GQ	0/1:255,0,255:99
52	plasmid1	66637	.	C	T	225	.	DP=71;VDB=0.0141;AF1=0.5;AC	GT:PL:GQ	0/1:255,0,249:99
53	plasmid1	66668	.	C	T	225	.	DP=80;VDB=0.0153;AF1=0.5;AC	GT:PL:GQ	0/1:255,0,255:99
54	plasmid2	2569	.	CAAAAA	CAAAA	177	.	INDEL;DP=110;VDB=0.0166;AF1	GT:PL:GQ	1/1:218,255,0:99
55	plasmid2	37805	.	TAAAAA	TAAAA	104	.	INDEL;DP=31;VDB=0.0058;AF1=	GT:PL:GQ	1/1:145,93,0:99

W18-3：INFO

	A	B	C	D	E	F	G	H
27	#CHROM	POS	ID	REF	ALT	QUAL	FIL	INFO
28	chromosome	141598	.	A	G	17	.	DP=7;VDB=0.0077;AF1=0.5;AC1=1;DP4=4,1,1,1;MQ=39;FQ=20.1;PV4=1,1,0,1
29	chromosome	242819	.	T	C	20	.	DP=9;VDB=0.0111;AF1=0.5;AC1=1;DP4=5,2,1,1;MQ=49;FQ=23;PV4=1,1,0,1
30	chromosome	359141	.	T	C	4.1	.	DP=14;VDB=0.0037;AF1=0.4998;AC1=1;DP4=11,1,1,1;MQ=49;FQ=6.2;PV4=0.27
31	chromosome	463882	.	T	G	6.2	.	DP=13;VDB=0.0074;AF1=0.4999;AC1=1;DP4=2,9,1,1;MQ=48;FQ=8.65;PV4=0.42
32	chromosome	660031	.	C	T	8.6	.	DP=11;VDB=0.0063;AF1=0.5;AC1=1;DP4=7,2,1,1;MQ=49;FQ=11.3;PV4=0.49,0.2
33	chromosome	792963	.	T	C	4.1	.	DP=14;VDB=0.0153;AF1=0.4998;AC1=1;DP4=2,10,1,1;MQ=48;FQ=6.2;PV4=0.4,1
34	chromosome	887422	.	G	A	10	.	DP=11;VDB=0.0070;AF1=0.5;AC1=1;DP4=7,2,1,1;MQ=49;FQ=13.2;PV4=0.49,1,0
35	chromosome	913470	.	T	C	14	.	DP=9;VDB=0.0056;AF1=0.5;AC1=1;DP4=1,6,1,1;MQ=49;FQ=17.1;PV4=0.42,1,0.2
36	chromosome	965171	.	T	C	17	.	DP=20;VDB=0.0071;AF1=0.5;AC1=1;DP4=14,3,2,1;MQ=49;FQ=20.1;PV4=0.51,1,
37	chromosome	1134950	.	A	G	7.8	.	DP=23;VDB=0.0076;AF1=0.5;AC1=1;DP4=5,15,1,2;MQ=49;FQ=10.4;PV4=1,1,0.0
38	chromosome	1160048	.	A	G	29	.	DP=17;VDB=0.0054;AF1=0.5;AC1=1;DP4=6,8,1,2;MQ=48;FQ=32;PV4=1,1,0.041,
39	chromosome	1160095	.	A	G	19	.	DP=10;VDB=0.0033;AF1=0.5;AC1=1;DP4=6,2,1,1;MQ=48;FQ=22;PV4=1,1,0.007,
40	chromosome	1240248	.	GCCCC	GCCCCC	214	.	INDEL;DP=48;VDB=0.0137;AF1=1;AC1=2;DP4=1,0,21,26;MQ=50;FQ=-163;PV4=
41	chromosome	1455552	.	CAAAA	CAAAAA	214	.	INDEL;DP=58;VDB=0.0164;AF1=1;AC1=2;DP4=2,1,28,27;MQ=49;FQ=-165;PV4=
42	chromosome	1475721	.	T	C	52	.	DP=19;VDB=0.0021;AF1=0.5;AC1=1;DP4=11,4,2,2;MQ=48;FQ=55;PV4=0.56,1,2.
43	chromosome	1752680	.	G	A	29	.	DP=24;VDB=0.0129;AF1=0.5;AC1=1;DP4=14,6,2,2;MQ=48;FQ=32;PV4=0.58,1,0.
44	chromosome	1941380	.	T	C	7.8	.	DP=13;VDB=0.0046;AF1=0.5;AC1=1;DP4=9,2,1,1;MQ=50;FQ=10.4;PV4=0.42,1,0
45	chromosome	2052955	.	A	G	12	.	DP=8;VDB=0.0084;AF1=0.5;AC1=1;DP4=2,3,2,1;MQ=50;FQ=15.1;PV4=1,0.34,0.1
46	chromosome	2054084	.	TC	T	14	.	INDEL;DP=8;VDB=0.0009;AF1=0.5;AC1=1;DP4=3,3,1,1;MQ=47;FQ=16.6;PV4=1,4
47	chromosome	2056279	.	A	G	24	.	DP=6;VDB=0.0057;AF1=0.5;AC1=1;DP4=2,2,1,1;MQ=46;FQ=27;PV4=1,1,1,1
48	chromosome	2057564	.	T	C	8.6	.	DP=10;VDB=0.0041;AF1=0.5;AC1=1;DP4=4,4,1,1;MQ=33;FQ=11.3;PV4=1,1,1,1
49	chromosome	2058145	.	C	T,G	108	.	DP=14;VDB=0.0147;AF1=1;AC1=2;DP4=0,0,7,7;MQ=23;FQ=-66
50	plasmid1	37177	.	G	T	54	.	DP=49;VDB=0.0087;AF1=0.5;AC1=1;DP4=24,16,5,4;MQ=49;FQ=57;PV4=1,0.1,0.
51	plasmid1	66609	.	A	G	225	.	DP=55;VDB=0.0060;AF1=0.5;AC1=1;DP4=16,8,22,9;MQ=43;FQ=225;PV4=0.78,1
52	plasmid1	66637	.	C	T	225	.	DP=71;VDB=0.0141;AF1=0.5;AC1=1;DP4=17,11,23,20;MQ=43;FQ=221;PV4=0.6
53	plasmid1	66668	.	C	T	225	.	DP=80;VDB=0.0153;AF1=0.5;AC1=1;DP4=17,17,23,23;MQ=43;FQ=225;PV4=1,0.
54	plasmid2	2569	.	CAAAAA	CAAAA	177	.	INDEL;DP=110;VDB=0.0166;AF1=1;AC1=2;DP4=0,0,51,59;MQ=49;FQ=-290
55	plasmid2	37805	.	TAAAAA	TAAAA	104	.	INDEL;DP=31;VDB=0.0058;AF1=1;AC1=2;DP4=0,0,5,26;MQ=31;FQ=-128

W18-3：INFO

①DPはdepthのこと。数が多いほど多くのリードがマップされたことを意味し、信頼度が高い

	A	B	C	D	E	F	G	H
27	#CHROM	POS	ID	REF	ALT	QUAL	FIL	INFO
28	chromosome	141598	.	A	G	17	.	DP=7;VDB=0.0077;AF1=0.5;AC1=1;DP4=4,1,1,1;MQ=39;FQ=20.1;PV4=1,1,0,1
29	chromosome	242819	.	T	C	20	.	DP=9;VDB=0.0111;AF1=0.5;AC1=1;DP4=5,2,1,1;MQ=49;FQ=23;PV4=1,1,0,1
30	chromosome	359141	.	T	C	4.1	.	DP=14;VDB=0.0037;AF1=0.4998;AC1=1;DP4=11,1,1,1;MQ=49;FQ=6.2;PV4=0.27
31	chromosome	463882	.	T	G	6.2	.	DP=13;VDB=0.0074;AF1=0.4999;AC1=1;DP4=2,9,1,1;MQ=48;FQ=8.65;PV4=0.42
32	chromosome	660031	.	C	T	8.6	.	DP=11;VDB=0.0063;AF1=0.5;AC1=1;DP4=7,2,1,1;MQ=49;FQ=11.3;PV4=0.49,0.2
33	chromosome	792963	.	T	C	4.1	.	DP=14;VDB=0.0153;AF1=0.4998;AC1=1;DP4=2,10,1,1;MQ=48;FQ=6.2;PV4=0.4,1
34	chromosome	887422	.	G	A	10	.	DP=11;VDB=0.0070;AF1=0.5;AC1=1;DP4=7,2,1,1;MQ=49;FQ=13.2;PV4=0.49,1,0
35	chromosome	913470	.	T	C	14	.	DP=9;VDB=0.0056;AF1=0.5;AC1=1;DP4=1,6,1,1;MQ=49;FQ=17.1;PV4=0.42,1,0.2
36	chromosome	965171	.	T	C	17	.	DP=20;VDB=0.0071;AF1=0.5;AC1=1;DP4=14,3,2,1;MQ=49;FQ=20.1;PV4=0.51,1,
37	chromosome	1134950	.	A	G	7.8	.	DP=23;VDB=0.0076;AF1=0.5;AC1=1;DP4=5,15,1,2;MQ=49;FQ=10.4;PV4=1,1,0.0
38	chromosome	1160048	.	A	G	29	.	DP=17;VDB=0.0054;AF1=0.5;AC1=1;DP4=6,8,1,2;MQ=48;FQ=32;PV4=1,1,0.041,
39	chromosome	1160095	.	A	G	19	.	DP=10;VDB=0.0033;AF1=0.5;AC1=1;DP4=6,2,1,1;MQ=48;FQ=22;PV4=1,1,0.007,
40	chromosome	1240248	.	GCCCC	GCCCCC	214	.	INDEL;DP=48;VDB=0.0137;AF1=1;AC1=2;DP4=1,0,21,26;MQ=50;FQ=-163;PV4=
41	chromosome	1455552	.	CAAAA	CAAAAA	214	.	INDEL;DP=58;VDB=0.0164;AF1=1;AC1=2;DP4=2,1,28,27;MQ=49;FQ=-165;PV4=
42	chromosome	1475721	.	T	C	52	.	DP=19;VDB=0.0021;AF1=0.5;AC1=1;DP4=11,4,2,2;MQ=48;FQ=55;PV4=0.56,1,2.
43	chromosome	1752680	.	G	A	29	.	DP=24;VDB=0.0129;AF1=0.5;AC1=1;DP4=14,6,2,2;MQ=48;FQ=32;PV4=0.58,1,0.
44	chromosome	1941380	.	T	C	7.8	.	DP=13;VDB=0.0046;AF1=0.5;AC1=1;DP4=9,2,1,1;MQ=50;FQ=10.4;PV4=0.42,1,0
45	chromosome	2052955	.	A	G	12	.	DP=8;VDB=0.0084;AF1=0.5;AC1=1;DP4=2,3,2,1;MQ=50;FQ=15.1;PV4=1,0.34,0.1
46	chromosome	2054084	.	TC	T	14	.	INDEL;DP=8;VDB=0.0009;AF1=0.5;AC1=1;DP4=3,3,1,1;MQ=47;FQ=16.6;PV4=1,4
47	chromosome	2056279	.	A	G	24	.	DP=6;VDB=0.0057;AF1=0.5;AC1=1;DP4=2,2,1,1;MQ=46;FQ=27;PV4=1,1,1,1
48	chromosome	2057564	.	T	C	8.6	.	DP=10;VDB=0.0041;AF1=0.5;AC1=1;DP4=4,4,1,1;MQ=33;FQ=11.3;PV4=1,1,1,1
49	chromosome	2058145	.	C	T,G	108	.	DP=14;VDB=0.0147;AF1=1;AC1=2;DP4=0,0,7,7;MQ=23;FQ=-66
50	plasmid1	37177	.	G	T	54	.	DP=49;VDB=0.0087;AF1=0.5;AC1=1;DP4=24,16,5,4;MQ=49;FQ=57;PV4=1,0.1,0.
51	plasmid1	66609	.	A	G	225	.	DP=55;VDB=0.0060;AF1=0.5;AC1=1;DP4=16,8,22,9;MQ=43;FQ=225;PV4=0.78,1
52	plasmid1	66637	.	C	T	225	.	DP=71;VDB=0.0141;AF1=0.5;AC1=1;DP4=17,11,23,20;MQ=43;FQ=221;PV4=0.6
53	plasmid1	66668	.	C	T	225	.	DP=80;VDB=0.0153;AF1=0.5;AC1=1;DP4=17,17,23,23;MQ=43;FQ=225;PV4=1,0.
54	plasmid2	2569	.	CAAAAA	CAAAA	177	.	INDEL;DP=110;VDB=0.0166;AF1=1;AC1=2;DP4=0,0,51,59;MQ=49;FQ=-290
55	plasmid2	37805	.	TAAAAA	TAAAA	104	.	INDEL;DP=31;VDB=0.0058;AF1=1;AC1=2;DP4=0,0,5,26;MQ=31;FQ=-128

W18-3：INFO

①VDBは、Variant Distance Biasの略。Higher values indicate higher likelihoods that the variant is distributed within the reads randomly.だそうです。(Biostars情報)。
実用上、ここを見て判断することはほぼない

	A	B	C	D	E	F	G	
27	#CHROM	POS	ID	REF	ALT	QUAL	FIL	INFO
28	chromosome	141598	.	A	G	17	.	DP=7;VDB=0.0077;AF1=0.5;AC1=1;DP4=4,1,1,1;MQ=39;FQ=20.1;PV4=1,1,0,1
29	chromosome	242819	.	T	C	20	.	DP=9;VDB=0.0111;AF1=0.5;AC1=1;DP4=5,2,1,1;MQ=49;FQ=23;PV4=1,1,0,1
30	chromosome	359141	.	T	C	4.1	.	DP=14;VDB=0.0037;AF1=0.4998;AC1=1;DP4=11,1,1,1;MQ=49;FQ=6.2;PV4=0.27
31	chromosome	463882	.	T	G	6.2	.	DP=13;VDB=0.0074;AF1=0.4999;AC1=1;DP4=2,9,1,1;MQ=48;FQ=8.65;PV4=0.42
32	chromosome	660031	.	C	T	8.6	.	DP=11;VDB=0.0063;AF1=0.5;AC1=1;DP4=7,2,1,1;MQ=49;FQ=11.3;PV4=0.49,0.2
33	chromosome	792963	.	T	C	4.1	.	DP=14;VDB=0.0153;AF1=0.4998;AC1=1;DP4=2,10,1,1;MQ=48;FQ=6.2;PV4=0.4,1
34	chromosome	887422	.	G	A	10	.	DP=11;VDB=0.0070;AF1=0.5;AC1=1;DP4=7,2,1,1;MQ=49;FQ=13.2;PV4=0.49,1,0
35	chromosome	913470	.	T	C	14	.	DP=9;VDB=0.0056;AF1=0.5;AC1=1;DP4=1,6,1,1;MQ=49;FQ=17.1;PV4=0.42,1,0.2
36	chromosome	965171	.	T	C	17	.	DP=20;VDB=0.0071;AF1=0.5;AC1=1;DP4=14,3,2,1;MQ=49;FQ=20.1;PV4=0.51,1,
37	chromosome	1134950	.	A	G	7.8	.	DP=23;VDB=0.0076;AF1=0.5;AC1=1;DP4=5,15,1,2;MQ=49;FQ=10.4;PV4=1,1,0.0
38	chromosome	1160048	.	A	G	29	.	DP=17;VDB=0.0054;AF1=0.5;AC1=1;DP4=6,8,1,2;MQ=48;FQ=32;PV4=1,1,0.041,
39	chromosome	1160095	.	A	G	19	.	DP=10;VDB=0.0033;AF1=0.5;AC1=1;DP4=6,2,1,1;MQ=48;FQ=22;PV4=1,1,0.007,
40	chromosome	1240248	.	GCCCC	GCCCCC	214	.	INDEL;DP=48;VDB=0.0137;AF1=1;AC1=2;DP4=1,0,21,26;MQ=50;FQ=-163;PV4=
41	chromosome	1455552	.	CAAAA	CAAAAA	214	.	INDEL;DP=58;VDB=0.0164;AF1=1;AC1=2;DP4=2,1,28,27;MQ=49;FQ=-165;PV4=
42	chromosome	1475721	.	T	C	52	.	DP=19;VDB=0.0021;AF1=0.5;AC1=1;DP4=11,4,2,2;MQ=48;FQ=55;PV4=0.56,1,2.
43	chromosome	1752680	.	G	A	29	.	DP=24;VDB=0.0129;AF1=0.5;AC1=1;DP4=14,6,2,2;MQ=48;FQ=32;PV4=0.58,1,0.
44	chromosome	1941380	.	T	C	7.8	.	DP=13;VDB=0.0046;AF1=0.5;AC1=1;DP4=9,2,1,1;MQ=50;FQ=10.4;PV4=0.42,1,0
45	chromosome	2052955	.	A	G	12	.	DP=8;VDB=0.0084;AF1=0.5;AC1=1;DP4=2,3,2,1;MQ=50;FQ=15.1;PV4=1,0.34,0.1
46	chromosome	2054084	.	TC	T	14	.	INDEL;DP=8;VDB=0.0009;AF1=0.5;AC1=1;DP4=3,3,1,1;MQ=47;FQ=16.6;PV4=1,4
47	chromosome	2056279	.	A	G	24	.	DP=6;VDB=0.0057;AF1=0.5;AC1=1;DP4=2,2,1,1;MQ=46;FQ=27;PV4=1,1,1,1
48	chromosome	2057564	.	T	C	8.6	.	DP=10;VDB=0.0041;AF1=0.5;AC1=1;DP4=4,4,1,1;MQ=33;FQ=11.3;PV4=1,1,1,1
49	chromosome	2058145	.	C	T,G	108	.	DP=14;VDB=0.0147;AF1=1;AC1=2;DP4=0,0,7,7;MQ=23;FQ=-66
50	plasmid1	37177	.	G	T	54	.	DP=49;VDB=0.0087;AF1=0.5;AC1=1;DP4=24,16,5,4;MQ=49;FQ=57;PV4=1,0.1,0.
51	plasmid1	66609	.	A	G	225	.	DP=55;VDB=0.0060;AF1=0.5;AC1=1;DP4=16,8,22,9;MQ=43;FQ=225;PV4=0.78,1
52	plasmid1	66637	.	C	T	225	.	DP=71;VDB=0.0141;AF1=0.5;AC1=1;DP4=17,11,23,20;MQ=43;FQ=221;PV4=0.6
53	plasmid1	66668	.	C	T	225	.	DP=80;VDB=0.0153;AF1=0.5;AC1=1;DP4=17,17,23,23;MQ=43;FQ=225;PV4=1,0.
54	plasmid2	2569	.	CAAAAA	CAAAA	177	.	INDEL;DP=110;VDB=0.0166;AF1=1;AC1=2;DP4=0,0,51,59;MQ=49;FQ=-290
55	plasmid2	37805	.	TAAAAA	TAAAA	104	.	INDEL;DP=31;VDB=0.0058;AF1=1;AC1=2;DP4=0,0,5,26;MQ=31;FQ=-128

W18-3：INFO

①AF1は、allele frequency（対立遺伝子頻度）のこと。よく見ると、最低でも0.4998よりも大きい値になっている。ここの値が約0.5以上のものがリストアップされているのだろう

	A	B	C	D	E	F	G	H
27	#CHROM	POS	ID	REF	ALT	QUAL	FIL	INFO
28	chromosome	141598	.	A	G	17	.	DP=7;VDB=0.0077;AF1=0.5;AC1=1;DP4=4,1,1,1;MQ=39;FQ=20.1;PV4=1,1,0,1
29	chromosome	242819	.	T	C	20	.	DP=9;VDB=0.0111;AF1=0.5;AC1=1;DP4=5,2,1,1;MQ=49;FQ=23;PV4=1,1,0,1
30	chromosome	359141	.	T	C	4.1	.	DP=14;VDB=0.0037;AF1=0.4998;AC1=1;DP4=11,1,1,1;MQ=49;FQ=6.2;PV4=0.27
31	chromosome	463882	.	T	G	6.2	.	DP=13;VDB=0.0074;AF1=0.4999;AC1=1;DP4=2,9,1,1;MQ=48;FQ=8.65;PV4=0.42
32	chromosome	660031	.	C	T	8.6	.	DP=11;VDB=0.0063;AF1=0.5;AC1=1;DP4=7,2,1,1;MQ=49;FQ=11.3;PV4=0.49,0.2
33	chromosome	792963	.	T	C	4.1	.	DP=14;VDB=0.0153;AF1=0.4998;AC1=1;DP4=2,10,1,1;MQ=48;FQ=6.2;PV4=0.4,1
34	chromosome	887422	.	G	A	10	.	DP=11;VDB=0.0070;AF1=0.5;AC1=1;DP4=7,2,1,1;MQ=49;FQ=13.2;PV4=0.49,1,0
35	chromosome	913470	.	T	C	14	.	DP=9;VDB=0.0056;AF1=0.5;AC1=1;DP4=1,6,1,1;MQ=49;FQ=17.1;PV4=0.42,1,0.2
36	chromosome	965171	.	T	C	17	.	DP=20;VDB=0.0071;AF1=0.5;AC1=1;DP4=14,3,2,1;MQ=49;FQ=20.1;PV4=0.51,1,
37	chromosome	1134950	.	A	G	7.8	.	DP=23;VDB=0.0076;AF1=0.5;AC1=1;DP4=5,15,1,2;MQ=49;FQ=10.4;PV4=1,1,0.0
38	chromosome	1160048	.	A	G	29	.	DP=17;VDB=0.0054;AF1=0.5;AC1=1;DP4=6,8,1,2;MQ=48;FQ=32;PV4=1,1,0.041,
39	chromosome	1160095	.	A	G	19	.	DP=10;VDB=0.0033;AF1=0.5;AC1=1;DP4=6,2,1,1;MQ=48;FQ=22;PV4=1,1,0.007,
40	chromosome	1240248	.	GCCCC	GCCCCC	214	.	INDEL;DP=48;VDB=0.0137;AF1=1;AC1=2;DP4=1,0,21,26;MQ=50;FQ=-163;PV4=
41	chromosome	1455552	.	CAAAA	CAAAAA	214	.	INDEL;DP=58;VDB=0.0164;AF1=1;AC1=2;DP4=2,1,28,27;MQ=49;FQ=-165;PV4=
42	chromosome	1475721	.	T	C	52	.	DP=19;VDB=0.0021;AF1=0.5;AC1=1;DP4=11,4,2,2;MQ=48;FQ=55;PV4=0.56,1,2.
43	chromosome	1752680	.	G	A	29	.	DP=24;VDB=0.0129;AF1=0.5;AC1=1;DP4=14,6,2,2;MQ=48;FQ=32;PV4=0.58,1,0.
44	chromosome	1941380	.	T	C	7.8	.	DP=13;VDB=0.0046;AF1=0.5;AC1=1;DP4=9,2,1,1;MQ=50;FQ=10.4;PV4=0.42,1,0
45	chromosome	2052955	.	A	G	12	.	DP=8;VDB=0.0084;AF1=0.5;AC1=1;DP4=2,3,2,1;MQ=50;FQ=15.1;PV4=1,0.34,0.1
46	chromosome	2054084	.	TC	T	14	.	INDEL;DP=8;VDB=0.0009;AF1=0.5;AC1=1;DP4=3,3,1,1;MQ=47;FQ=16.6;PV4=1,4
47	chromosome	2056279	.	A	G	24	.	DP=6;VDB=0.0057;AF1=0.5;AC1=1;DP4=2,2,1,1;MQ=46;FQ=27;PV4=1,1,1,1
48	chromosome	2057564	.	T	C	8.6	.	DP=10;VDB=0.0041;AF1=0.5;AC1=1;DP4=4,4,1,1;MQ=33;FQ=11.3;PV4=1,1,1,1
49	chromosome	2058145	.	C	T,G	108	.	DP=14;VDB=0.0147;AF1=1;AC1=2;DP4=0,0,7,7;MQ=23;FQ=-66
50	plasmid1	37177	.	G	T	54	.	DP=49;VDB=0.0087;AF1=0.5;AC1=1;DP4=24,16,5,4;MQ=49;FQ=57;PV4=1,0.1,0.
51	plasmid1	66609	.	A	G	225	.	DP=55;VDB=0.0060;AF1=0.5;AC1=1;DP4=16,8,22,9;MQ=43;FQ=225;PV4=0.78,1
52	plasmid1	66637	.	C	T	225	.	DP=71;VDB=0.0141;AF1=0.5;AC1=1;DP4=17,11,23,20;MQ=43;FQ=221;PV4=0.6
53	plasmid1	66668	.	C	T	225	.	DP=80;VDB=0.0153;AF1=0.5;AC1=1;DP4=17,17,23,23;MQ=43;FQ=225;PV4=1,0.
54	plasmid2	2569	.	CAAAAA	CAAAA	177	.	INDEL;DP=110;VDB=0.0166;AF1=1;AC1=2;DP4=0,0,51,59;MQ=49;FQ=-290
55	plasmid2	37805	.	TAAAAA	TAAAA	104	.	INDEL;DP=31;VDB=0.0058;AF1=1;AC1=2;DP4=0,0,5,26;MQ=31;FQ=-128

W18-3：INFO

①AC1は、genotype中のallele count（対立遺伝子数）のこと。ほとんどがAC1=1だが、赤矢印で示す5個がAC1=2になっていることがわかる

	A	B	C	D	E	F	G	H
27	#CHROM	POS	ID	REF	ALT	QUAL	FIL	INFO
28	chromosome	141598	.	A	G	17	.	DP=7;VDB=0.0077;AF1=0.5;AC1=1;DP4=4,1,1,1;MQ=39;FQ=20.1;PV4=1,1,0,1
29	chromosome	242819	.	T	C	20	.	DP=9;VDB=0.0111;AF1=0.5;AC1=1;DP4=5,2,1,1;MQ=49;FQ=23;PV4=1,1,0,1
30	chromosome	359141	.	T	C	4.1	.	DP=14;VDB=0.0037;AF1=0.4998;AC1=1;DP4=11,1,1,1;MQ=49;FQ=6.2;PV4=0.27
31	chromosome	463882	.	T	G	6.2	.	DP=13;VDB=0.0074;AF1=0.4999;AC1=1;DP4=2,9,1,1;MQ=48;FQ=8.65;PV4=0.42
32	chromosome	660031	.	C	T	8.6	.	DP=11;VDB=0.0063;AF1=0.5;AC1=1;DP4=7,2,1,1;MQ=49;FQ=11.3;PV4=0.49,0.2
33	chromosome	792963	.	T	C	4.1	.	DP=14;VDB=0.0153;AF1=0.4998;AC1=1;DP4=2,10,1,1;MQ=48;FQ=6.2;PV4=0.4,1
34	chromosome	887422	.	G	A	10	.	DP=11;VDB=0.0070;AF1=0.5;AC1=1;DP4=7,2,1,1;MQ=49;FQ=13.2;PV4=0.49,1,0
35	chromosome	913470	.	T	C	14	.	DP=9;VDB=0.0056;AF1=0.5;AC1=1;DP4=1,6,1,1;MQ=49;FQ=17.1;PV4=0.42,1,0.2
36	chromosome	965171	.	T	C	17	.	DP=20;VDB=0.0071;AF1=0.5;AC1=1;DP4=14,3,2,1;MQ=49;FQ=20.1;PV4=0.51,1,
37	chromosome	1134950	.	A	G	7.8	.	DP=23;VDB=0.0076;AF1=0.5;AC1=1;DP4=5,15,1,2;MQ=49;FQ=10.4;PV4=1,1,0.0
38	chromosome	1160048	.	A	G	29	.	DP=17;VDB=0.0054;AF1=0.5;AC1=1;DP4=6,8,1,2;MQ=48;FQ=32;PV4=1,1,0.041,
39	chromosome	1160095	.	A	G	19	.	DP=10;VDB=0.0033;AF1=0.5;AC1=1;DP4=6,2,1,1;MQ=48;FQ=22;PV4=1,1,0.007,
40	chromosome	1240248	.	GCCCC	GCCCCC	214	.	INDEL;DP=48;VDB=0.0137;AF1=1;AC1=2;DP4=1,0,21,26;MQ=50;FQ=-163;PV4=
41	chromosome	1455552	.	CAAAA	CAAAAA	214	.	INDEL;DP=58;VDB=0.0164;AF1=1;AC1=2;DP4=2,1,28,27;MQ=49;FQ=-165;PV4=
42	chromosome	1475721	.	T	C	52	.	DP=19;VDB=0.0021;AF1=0.5;AC1=1;DP4=11,4,2,2;MQ=48;FQ=55;PV4=0.56,1,2.
43	chromosome	1752680	.	G	A	29	.	DP=24;VDB=0.0129;AF1=0.5;AC1=1;DP4=14,6,2,2;MQ=48;FQ=32;PV4=0.58,1,0.
44	chromosome	1941380	.	T	C	7.8	.	DP=13;VDB=0.0046;AF1=0.5;AC1=1;DP4=9,2,1,1;MQ=50;FQ=10.4;PV4=0.42,1,0
45	chromosome	2052955	.	A	G	12	.	DP=8;VDB=0.0084;AF1=0.5;AC1=1;DP4=2,3,2,1;MQ=50;FQ=15.1;PV4=1,0.34,0.1
46	chromosome	2054084	.	TC	T	14	.	INDEL;DP=8;VDB=0.0009;AF1=0.5;AC1=1;DP4=3,3,1,1;MQ=47;FQ=16.6;PV4=1,4
47	chromosome	2056279	.	A	G	24	.	DP=6;VDB=0.0057;AF1=0.5;AC1=1;DP4=2,2,1,1;MQ=46;FQ=27;PV4=1,1,1,1
48	chromosome	2057564	.	T	C	8.6	.	DP=10;VDB=0.0041;AF1=0.5;AC1=1;DP4=4,4,1,1;MQ=33;FQ=11.3;PV4=1,1,1,1
49	chromosome	2058145	.	C	T,G	108	.	DP=14;VDB=0.0147;AF1=1;AC1=2;DP4=0,0,7,7;MQ=23;FQ=-66
50	plasmid1	37177	.	G	T	54	.	DP=49;VDB=0.0087;AF1=0.5;AC1=1;DP4=24,16,5,4;MQ=49;FQ=57;PV4=1,0.1,0.
51	plasmid1	66609	.	A	G	225	.	DP=55;VDB=0.0060;AF1=0.5;AC1=1;DP4=16,8,22,9;MQ=43;FQ=225;PV4=0.78,1
52	plasmid1	66637	.	C	T	225	.	DP=71;VDB=0.0141;AF1=0.5;AC1=1;DP4=17,11,23,20;MQ=43;FQ=221;PV4=0.6
53	plasmid1	66668	.	C	T	225	.	DP=80;VDB=0.0153;AF1=0.5;AC1=1;DP4=17,17,23,23;MQ=43;FQ=225;PV4=1,0.
54	plasmid2	2569	.	CAAAAA	CAAAA	177	.	INDEL;DP=110;VDB=0.0166;AF1=1;AC1=2;DP4=0,0,51,59;MQ=49;FQ=-290
55	plasmid2	37805	.	TAAAAA	TAAAA	104	.	INDEL;DP=31;VDB=0.0058;AF1=1;AC1=2;DP4=0,0,5,26;MQ=31;FQ=-128

W18-3：INFO

①DP4は、1番左のDPを割り振ったもの。DP4=REFのforward,REFのreverse,ALTのforward,ALTのreverseの順番

	A	B	C	D	E	F	G	H
27	#CHROM	POS	ID	REF	ALT	QUAL	FIL	INFO
28	chromosome	141598	.	A	G	17	.	DP=7;VDB=0.0077;AF1=0.5;AC1=1;DP4=4,1,1,1;MQ=39;FQ=20.1;PV4=1,1,0,1
29	chromosome	242819	.	T	C	20	.	DP=9;VDB=0.0111;AF1=0.5;AC1=1;DP4=5,2,1,1;MQ=49;FQ=23;PV4=1,1,0,1
30	chromosome	359141	.	T	C	4.1	.	DP=14;VDB=0.0037;AF1=0.4998;AC1=1;DP4=11,1,1,1;MQ=49;FQ=6.2;PV4=0.27
31	chromosome	463882	.	T	G	6.2	.	DP=13;VDB=0.0074;AF1=0.4999;AC1=1;DP4=2,9,1,1;MQ=48;FQ=8.65;PV4=0.42
32	chromosome	660031	.	C	T	8.6	.	DP=11;VDB=0.0063;AF1=0.5;AC1=1;DP4=7,2,1,1;MQ=49;FQ=11.3;PV4=0.49,0.2
33	chromosome	792963	.	T	C	4.1	.	DP=14;VDB=0.0153;AF1=0.4998;AC1=1;DP4=2,10,1,1;MQ=48;FQ=6.2;PV4=0.4,1
34	chromosome	887422	.	G	A	10	.	DP=11;VDB=0.0070;AF1=0.5;AC1=1;DP4=7,2,1,1;MQ=49;FQ=13.2;PV4=0.49,1,0
35	chromosome	913470	.	T	C	14	.	DP=9;VDB=0.0056;AF1=0.5;AC1=1;DP4=1,6,1,1;MQ=49;FQ=17.1;PV4=0.42,1,0.2
36	chromosome	965171	.	T	C	17	.	DP=20;VDB=0.0071;AF1=0.5;AC1=1;DP4=14,3,2,1;MQ=49;FQ=20.1;PV4=0.51,1,
37	chromosome	1134950	.	A	G	7.8	.	DP=23;VDB=0.0076;AF1=0.5;AC1=1;DP4=5,15,1,2;MQ=49;FQ=10.4;PV4=1,1,0.0
38	chromosome	1160048	.	A	G	29	.	DP=17;VDB=0.0054;AF1=0.5;AC1=1;DP4=6,8,1,2;MQ=48;FQ=32;PV4=1,1,0.041,
39	chromosome	1160095	.	A	G	19	.	DP=10;VDB=0.0033;AF1=0.5;AC1=1;DP4=6,2,1,1;MQ=48;FQ=22;PV4=1,1,0.007,
40	chromosome	1240248	.	GCCCC	GCCCCC	214	.	INDEL;DP=48;VDB=0.0137;AF1=1;AC1=2;DP4=1,0,21,26;MQ=50;FQ=-163;PV4=
41	chromosome	1455552	.	CAAAA	CAAAAA	214	.	INDEL;DP=58;VDB=0.0164;AF1=1;AC1=2;DP4=2,1,28,27;MQ=49;FQ=-165;PV4=
42	chromosome	1475721	.	T	C	52	.	DP=19;VDB=0.0021;AF1=0.5;AC1=1;DP4=11,4,2,2;MQ=48;FQ=55;PV4=0.56,1,2.
43	chromosome	1752680	.	G	A	29	.	DP=24;VDB=0.0129;AF1=0.5;AC1=1;DP4=14,6,2,2;MQ=48;FQ=32;PV4=0.58,1,0.
44	chromosome	1941380	.	T	C	7.8	.	DP=13;VDB=0.0046;AF1=0.5;AC1=1;DP4=9,2,1,1;MQ=50;FQ=10.4;PV4=0.42,1,0
45	chromosome	2052955	.	A	G	12	.	DP=8;VDB=0.0084;AF1=0.5;AC1=1;DP4=2,3,2,1;MQ=50;FQ=15.1;PV4=1,0.34,0.1
46	chromosome	2054084	.	TC	T	14	.	INDEL;DP=8;VDB=0.0009;AF1=0.5;AC1=1;DP4=3,3,1,1;MQ=47;FQ=16.6;PV4=1,4
47	chromosome	2056279	.	A	G	24	.	DP=6;VDB=0.0057;AF1=0.5;AC1=1;DP4=2,2,1,1;MQ=46;FQ=27;PV4=1,1,1,1
48	chromosome	2057564	.	T	C	8.6	.	DP=10;VDB=0.0041;AF1=0.5;AC1=1;DP4=4,4,1,1;MQ=33;FQ=11.3;PV4=1,1,1,1
49	chromosome	2058145	.	C	T,G	108	.	DP=14;VDB=0.0147;AF1=1;AC1=2;DP4=0,0,7,7;MQ=23;FQ=-66
50	plasmid1	37177	.	G	T	54	.	DP=49;VDB=0.0087;AF1=0.5;AC1=1;DP4=24,16,5,4;MQ=49;FQ=57;PV4=1,0.1,0.
51	plasmid1	66609	.	A	G	225	.	DP=55;VDB=0.0060;AF1=0.5;AC1=1;DP4=16,8,22,9;MQ=43;FQ=225;PV4=0.78,1
52	plasmid1	66637	.	C	T	225	.	DP=71;VDB=0.0141;AF1=0.5;AC1=1;DP4=17,11,23,20;MQ=43;FQ=221;PV4=0.6
53	plasmid1	66668	.	C	T	225	.	DP=80;VDB=0.0153;AF1=0.5;AC1=1;DP4=17,17,23,23;MQ=43;FQ=225;PV4=1,0.
54	plasmid2	2569	.	CAAAAA	CAAAA	177	.	INDEL;DP=110;VDB=0.0166;AF1=1;AC1=2;DP4=0,0,51,59;MQ=49;FQ=-290
55	plasmid2	37805	.	TAAAAA	TAAAA	104	.	INDEL;DP=31;VDB=0.0058;AF1=1;AC1=2;DP4=0,0,5,26;MQ=31;FQ=-128

W18-3：INF

例えば①「DP4=4,1,1,1」は②DP=7を割り振ったもの。概要配列と同じ塩基（REF；つまり③A）のforward鎖側でマップされたのが4リード（depth=4）、REFのreverse側がdepth=1、変異塩基（ALT；つまり④G）でforward側がdepth=1、変異塩基（ALT；つまり④G）でreverse側がdepth=1

	A	B	C	D	E	F	G	H
27	#CHROM	POS	ID	REF	ALT	QUAL	FILTER	INFO
28	chromosome	141598	.	A	G	17	.	DP=7;VDB=0.0077;AF1=0.5;AC1=1;DP4=4,1,1,1;MQ=39;FQ=20.1;PV4=1,1,0,1
29	chromosome	242819	.			20	.	VDB=0.0111;AF1=0.5;AC1=1;DP4=1,1,1,1;MQ=49;FQ=23;PV4=1,1,0,1
30	chromosome	359141	.	T	C	4.1	.	DP=14;VDB=0.0037;AF1=0.4998;AC1=1;DP4=11,1,1,1;MQ=49;FQ=6.2;PV4=0.27
31	chromosome	463882	.	T	G	6.2	.	DP=13;VDB=0.0074;AF1=0.4999;AC1=1;DP4=2,9,1,1;MQ=48;FQ=8.65;PV4=0.42
32	chromosome	660031	.	C	T	8.6	.	DP=11;VDB=0.0063;AF1=0.5;AC1=1;DP4=7,2,1,1;MQ=49;FQ=11.3;PV4=0.49,0.2
33	chromosome	792963	.	T	C	4.1	.	DP=14;VDB=0.0153;AF1=0.4998;AC1=1;DP4=2,10,1,1;MQ=48;FQ=6.2;PV4=0.4,1
34	chromosome	887422	.	G	A	10	.	DP=11;VDB=0.0070;AF1=0.5;AC1=1;DP4=7,2,1,1;MQ=49;FQ=13.2;PV4=0.49,1,0
35	chromosome	913470	.	T	C	14	.	DP=9;VDB=0.0056;AF1=0.5;AC1=1;DP4=1,6,1,1;MQ=49;FQ=17.1;PV4=0.42,1,0.2
36	chromosome	965171	.	T	C	17	.	DP=20;VDB=0.0071;AF1=0.5;AC1=1;DP4=14,3,2,1;MQ=49;FQ=20.1;PV4=0.51,1,
37	chromosome	1134950	.	A	G	7.8	.	DP=23;VDB=0.0076;AF1=0.5;AC1=1;DP4=5,15,1,2;MQ=49;FQ=10.4;PV4=1,1,0.0
38	chromosome	1160048	.	A	G	29	.	DP=17;VDB=0.0054;AF1=0.5;AC1=1;DP4=6,8,1,2;MQ=48;FQ=32;PV4=1,1,0.041,
39	chromosome	1160095	.	A	G	19	.	DP=10;VDB=0.0033;AF1=0.5;AC1=1;DP4=6,2,1,1;MQ=48;FQ=22;PV4=1,1,0.007,
40	chromosome	1240248	.	GCCCC	GCCCCC	214	.	INDEL;DP=48;VDB=0.0137;AF1=1;AC1=2;DP4=1,0,21,26;MQ=50;FQ=-163;PV4=
41	chromosome	1455552	.	CAAAA	CAAAAA	214	.	INDEL;DP=58;VDB=0.0164;AF1=1;AC1=2;DP4=2,1,28,27;MQ=49;FQ=-165;PV4=
42	chromosome	1475721	.	T	C	52	.	DP=19;VDB=0.0021;AF1=0.5;AC1=1;DP4=11,4,2,2;MQ=48;FQ=55;PV4=0.56,1,2.
43	chromosome	1752680	.	G	A	29	.	DP=24;VDB=0.0129;AF1=0.5;AC1=1;DP4=14,6,2,2;MQ=48;FQ=32;PV4=0.58,1,0.
44	chromosome	1941380	.	T	C	7.8	.	DP=13;VDB=0.0046;AF1=0.5;AC1=1;DP4=9,2,1,1;MQ=50;FQ=10.4;PV4=0.42,1,0
45	chromosome	2052955	.	A	G	12	.	DP=8;VDB=0.0084;AF1=0.5;AC1=1;DP4=2,3,2,1;MQ=50;FQ=15.1;PV4=1,0.34,0.1
46	chromosome	2054084	.	TC	T	14	.	INDEL;DP=8;VDB=0.0009;AF1=0.5;AC1=1;DP4=3,3,1,1;MQ=47;FQ=16.6;PV4=1,4
47	chromosome	2056279	.	A	G	24	.	DP=6;VDB=0.0057;AF1=0.5;AC1=1;DP4=2,2,1,1;MQ=46;FQ=27;PV4=1,1,1,1
48	chromosome	2057564	.	T	C	8.6	.	DP=10;VDB=0.0041;AF1=0.5;AC1=1;DP4=4,4,1,1;MQ=33;FQ=11.3;PV4=1,1,1,1
49	chromosome	2058145	.	C	T,G	108	.	DP=14;VDB=0.0147;AF1=1;AC1=2;DP4=0,0,7,7;MQ=23;FQ=-66
50	plasmid1	37177	.	G	T	54	.	DP=49;VDB=0.0087;AF1=0.5;AC1=1;DP4=24,16,5,4;MQ=49;FQ=57;PV4=1,0.1,0.
51	plasmid1	66609	.	A	G	225	.	DP=55;VDB=0.0060;AF1=0.5;AC1=1;DP4=16,8,22,9;MQ=43;FQ=225;PV4=0.78,1
52	plasmid1	66637	.	C	T	225	.	DP=71;VDB=0.0141;AF1=0.5;AC1=1;DP4=17,11,23,20;MQ=43;FQ=221;PV4=0.6
53	plasmid1	66668	.	C	T	225	.	DP=80;VDB=0.0153;AF1=0.5;AC1=1;DP4=17,17,23,23;MQ=43;FQ=225;PV4=1,0.
54	plasmid2	2569	.	CAAAAA	CAAAA	177	.	INDEL;DP=110;VDB=0.0166;AF1=1;AC1=2;DP4=0,0,51,59;MQ=49;FQ=-290
55	plasmid2	37805	.	TAAAAA	TAAAA	104	.	INDEL;DP=31;VDB=0.0058;AF1=1;AC1=2;DP4=0,0,5,26;MQ=31;FQ=-128

W18-3：INFO

概要配列を変更するかどうかの判断基準の1つとして、①を眺める。この場合は、②REFのAを支持するリード数が4+1=5、③ALTのGを支持するリード数が1+1=2。したがって、概要配列であるREF側のAをGには変更しないという結論を下す

	A	B	C	D	E	F		
27	#CHROM	POS	ID	REF	ALT	QUAL	FILTER	INFO
28	chromosome	141598	.	A	G	17	.	DP=7;VDB=0.0077;AF1=0.5;AC1=1;DP4=4,1,1,1;MQ=39;FQ=20.1;PV4=1,1,0,1
29	chromosome	242819	.	②	③	20	.	DP=9;VDB=0.0111;AF1=0.5;AC1=1;DP4=3,1,1,1;MQ=49;FQ=23;PV4=1,1,0,1
30	chromosome	359141	.	T	C	4.1	.	DP=14;VDB=0.0037;AF1=0.4998;AC1=1;DP4=11,1,1,1;MQ=49;FQ=6.2;PV4=0.27
31	chromosome	463882	.	T	G	6.2	.	DP=13;VDB=0.0074;AF1=0.4999;AC1=1;DP4=2,9,1,1;MQ=48;FQ=8.65;PV4=0.42
32	chromosome	660031	.	C	T	8.6	.	DP=11;VDB=0.0063;AF1=0.5;AC1=1;DP4=7,2,1,1;MQ=49;FQ=11.3;PV4=0.49,0.2
33	chromosome	792963	.	T	C	4.1	.	DP=14;VDB=0.0153;AF1=0.4998;AC1=1;DP4=2,10,1,1;MQ=48;FQ=6.2;PV4=0.4,1
34	chromosome	887422	.	G	A	10	.	DP=11;VDB=0.0070;AF1=0.5;AC1=1;DP4=7,2,1,1;MQ=49;FQ=13.2;PV4=0.49,1,0
35	chromosome	913470	.	T	C	14	.	DP=9;VDB=0.0056;AF1=0.5;AC1=1;DP4=1,6,1,1;MQ=49;FQ=17.1;PV4=0.42,1,0.2
36	chromosome	965171	.	T	C	17	.	DP=20;VDB=0.0071;AF1=0.5;AC1=1;DP4=14,3,2,1;MQ=49;FQ=20.1;PV4=0.51,1,
37	chromosome	1134950	.	A	G	7.8	.	DP=23;VDB=0.0076;AF1=0.5;AC1=1;DP4=5,15,1,2;MQ=49;FQ=10.4;PV4=1,1,0.0
38	chromosome	1160048	.	A	G	29	.	DP=17;VDB=0.0054;AF1=0.5;AC1=1;DP4=6,8,1,2;MQ=48;FQ=32;PV4=1,1,0.041,
39	chromosome	1160095	.	A	G	19	.	DP=10;VDB=0.0033;AF1=0.5;AC1=1;DP4=6,2,1,1;MQ=48;FQ=22;PV4=1,1,0.007,
40	chromosome	1240248	.	GCCCC	GCCCCC	214	.	INDEL;DP=48;VDB=0.0137;AF1=1;AC1=2;DP4=1,0,21,26;MQ=50;FQ=-163;PV4=
41	chromosome	1455552	.	CAAAA	CAAAAA	214	.	INDEL;DP=58;VDB=0.0164;AF1=1;AC1=2;DP4=2,1,28,27;MQ=49;FQ=-165;PV4=
42	chromosome	1475721	.	T	C	52	.	DP=19;VDB=0.0021;AF1=0.5;AC1=1;DP4=11,4,2,2;MQ=48;FQ=55;PV4=0.56,1,2.
43	chromosome	1752680	.	G	A	29	.	DP=24;VDB=0.0129;AF1=0.5;AC1=1;DP4=14,6,2,2;MQ=48;FQ=32;PV4=0.58,1,0.
44	chromosome	1941380	.	T	C	7.8	.	DP=13;VDB=0.0046;AF1=0.5;AC1=1;DP4=9,2,1,1;MQ=50;FQ=10.4;PV4=0.42,1,0
45	chromosome	2052955	.	A	G	12	.	DP=8;VDB=0.0084;AF1=0.5;AC1=1;DP4=2,3,2,1;MQ=50;FQ=15.1;PV4=1,0.34,0.1
46	chromosome	2054084	.	TC	T	14	.	INDEL;DP=8;VDB=0.0009;AF1=0.5;AC1=1;DP4=3,3,1,1;MQ=47;FQ=16.6;PV4=1,4
47	chromosome	2056279	.	A	G	24	.	DP=6;VDB=0.0057;AF1=0.5;AC1=1;DP4=2,2,1,1;MQ=46;FQ=27;PV4=1,1,1,1
48	chromosome	2057564	.	T	C	8.6	.	DP=10;VDB=0.0041;AF1=0.5;AC1=1;DP4=4,4,1,1;MQ=33;FQ=11.3;PV4=1,1,1,1
49	chromosome	2058145	.	C	T,G	108	.	DP=14;VDB=0.0147;AF1=1;AC1=2;DP4=0,0,7,7;MQ=23;FQ=-66
50	plasmid1	37177	.	G	T	54	.	DP=49;VDB=0.0087;AF1=0.5;AC1=1;DP4=24,16,5,4;MQ=49;FQ=57;PV4=1,0.1,0.
51	plasmid1	66609	.	A	G	225	.	DP=55;VDB=0.0060;AF1=0.5;AC1=1;DP4=16,8,22,9;MQ=43;FQ=225;PV4=0.78,1
52	plasmid1	66637	.	C	T	225	.	DP=71;VDB=0.0141;AF1=0.5;AC1=1;DP4=17,11,23,20;MQ=43;FQ=221;PV4=0.6
53	plasmid1	66668	.	C	T	225	.	DP=80;VDB=0.0153;AF1=0.5;AC1=1;DP4=17,17,23,23;MQ=43;FQ=225;PV4=1,0.
54	plasmid2	2569	.	CAAAAA	CAAAA	177	.	INDEL;DP=110;VDB=0.0166;AF1=1;AC1=2;DP4=0,0,51,59;MQ=49;FQ=-290
55	plasmid2	37805	.	TAAAAA	TAAAA	104	.	INDEL;DP=31;VDB=0.0058;AF1=1;AC1=2;DP4=0,0,5,26;MQ=31;FQ=-128

W18-3：INFO

概要配列を変更するかどうかの判断基準の1つとして、①を眺める。この場合は、②REFのGCCCCを支持するリード数が1+0=1、③ALTのGCCCCCを支持するリード数が21+26=47。したがって、概要配列であるREF側のGCCCCをGCCCCCに変更したほうがよいと判断する

	A	B	C	D				
27	#CHROM	POS	ID	REF	ALT	QUAL	FILTER	INFO
28	chromosome	141598	.	A	G	17	.	DP=7;VDB=0.0077;AF1=0.5;AC1=1;DP4=4,1,1,1;MQ=39;FQ=20.1;PV4=1,1,0,1
29	chromosome	242819	.	T	C	20	.	DP=9;VDB=0.0111;AF1=0.5;AC1=1;DP4=5,2,1,1;MQ=49;FQ=23;PV4=1,1,0,1
30	chromosome	359141	.	T	C	4.1	.	DP=14;VDB=0.0037;AF1=0.4998;AC1=1;DP4=11,1,1,1;MQ=49;FQ=6.2;PV4=0.27
31	chromosome	463882	.	T	G	6.2	.	DP=13;VDB=0.0074;AF1=0.4999;AC1=1;DP4=2,9,1,1;MQ=48;FQ=8.65;PV4=0.42
32	chromosome	660031	.	C	T	8.6	.	DP=11;VDB=0.0063;AF1=0.5;AC1=1;DP4=7,2,1,1;MQ=49;FQ=11.3;PV4=0.49,0.2
33	chromosome	792963	.	T	C	4.1	.	DP=14;VDB=0.0153;AF1=0.4998;AC1=1;DP4=2,10,1,1;MQ=48;FQ=6.2;PV4=0.4,1
34	chromosome	887422	.	G	A	10	.	DP=11;VDB=0.0070;AF1=0.5;AC1=1;DP4=7,2,1,1;MQ=49;FQ=13.2;PV4=0.49,1,0
35	chromosome	913470	.	T	C	14	.	DP=9;VDB=0.0056;AF1=0.5;AC1=1;DP4=1,6,1,1;MQ=49;FQ=17.1;PV4=0.42,1,0.2
36	chromosome	965171	.	T	C	17	.	DP=20;VDB=0.0071;AF1=0.5;AC1=1;DP4=14,3,2,1;MQ=49;FQ=20.1;PV4=0.51,1,
37	chromosome	1134950	.	A	G	7.8	.	DP=23;VDB=0.0076;AF1=0.5;AC1=1;DP4=5,15,1,2;MQ=49;FQ=10.4;PV4=1,1,0.0
38	chromosome	1160048	.	A	G	29	.	DP=17;VDB=0.0054;AF1=0.5;AC1=1;DP4=6,8,1,2;MQ=48;FQ=32;PV4=1,1,0.041,
39	chromosome	1160095	.	A	G	19	.	DP=10;VDB=0.0033;AF1=0.5;AC1=1;DP4=6,2,1,1;MQ=48;FQ=22;PV4=1,1,0.007,
40	chromosome	1240248	.	GCCCC	GCCCCC	214	.	INDEL;DP=48;VDB=0.0137;AF1=1;AC1=2;DP4=1,0,21,26;MQ=50;FQ=-163;PV4=
41	chromosome	1455552	.	AAAAA	AAAAAA	214	.	INDEL;DP=58;VDB=0.0164;AF1=1;AC1=2;DP4=[illegible]28,27;MQ=49;FQ=-165;PV4=
42	chromosome	1475721	.	T	C	52	.	DP=19;VDB=0.0021;AF1=0.5;AC1=1;DP4=11,4,2,2;MQ=48;FQ=55;PV4=0.56,1,2.
43	chromosome	1752680	.	G	A	29	.	DP=24;VDB=0.0129;AF1=0.5;AC1=1;DP4=14,6,2,2;MQ=48;FQ=32;PV4=0.58,1,0.
44	chromosome	1941380	.	T	C	7.8	.	DP=13;VDB=0.0046;AF1=0.5;AC1=1;DP4=9,2,1,1;MQ=50;FQ=10.4;PV4=0.42,1,0
45	chromosome	2052955	.	A	G	12	.	DP=8;VDB=0.0084;AF1=0.5;AC1=1;DP4=2,3,2,1;MQ=50;FQ=15.1;PV4=1,0.34,0.1
46	chromosome	2054084	.	TC	T	14	.	INDEL;DP=8;VDB=0.0009;AF1=0.5;AC1=1;DP4=3,3,1,1;MQ=47;FQ=16.6;PV4=1,4
47	chromosome	2056279	.	A	G	24	.	DP=6;VDB=0.0057;AF1=0.5;AC1=1;DP4=2,2,1,1;MQ=46;FQ=27;PV4=1,1,1,1
48	chromosome	2057564	.	T	C	8.6	.	DP=10;VDB=0.0041;AF1=0.5;AC1=1;DP4=4,4,1,1;MQ=33;FQ=11.3;PV4=1,1,1,1
49	chromosome	2058145	.	C	T,G	108	.	DP=14;VDB=0.0147;AF1=1;AC1=2;DP4=0,0,7,7;MQ=23;FQ=-66
50	plasmid1	37177	.	G	T	54	.	DP=49;VDB=0.0087;AF1=0.5;AC1=1;DP4=24,16,5,4;MQ=49;FQ=57;PV4=1,0.1,0.
51	plasmid1	66609	.	A	G	225	.	DP=55;VDB=0.0060;AF1=0.5;AC1=1;DP4=16,8,22,9;MQ=43;FQ=225;PV4=0.78,1
52	plasmid1	66637	.	C	T	225	.	DP=71;VDB=0.0141;AF1=0.5;AC1=1;DP4=17,11,23,20;MQ=43;FQ=221;PV4=0.6
53	plasmid1	66668	.	C	T	225	.	DP=80;VDB=0.0153;AF1=0.5;AC1=1;DP4=17,17,23,23;MQ=43;FQ=225;PV4=1,0.
54	plasmid2	2569	.	CAAAAA	CAAAA	177	.	INDEL;DP=110;VDB=0.0166;AF1=1;AC1=2;DP4=0,0,51,59;MQ=49;FQ=-290
55	plasmid2	37805	.	TAAAAA	TAAAA	104	.	INDEL;DP=31;VDB=0.0058;AF1=1;AC1=2;DP4=0,0,5,26;MQ=31;FQ=-128

W18-4：全体

もう一度全体を眺め直す。VCFファイル中にリストアップされているのは、概要配列と異なる部分のみ。概要配列と同じ塩基がマップされた箇所は、①がおそらく0/0と表現されるが、ないので妥当

	A	B	C	D	E	F	G	H	I	J
27	#CHROM	POS	ID	REF	ALT	QUAL	FILTER	INFO	FORMAT	①me/w3pipeline/refdata/
28	chromosome	141598	.	A	G	17.1	.	DP=7;VDB=0.0077;AF1=0.5;AC1	GT:PL:GQ	0/1:47,0,138:50
29	chromosome	242819	.	T	C	20	.	DP=9;VDB=0.0111;AF1=0.5;AC1	GT:PL:GQ	0/1:50,0,139:53
30	chromosome	359141	.	T	C	4.13	.	DP=14;VDB=0.0037;AF1=0.4998	GT:PL:GQ	0/1:32,0,187:31
31	chromosome	463882	.	T	G	6.2	.	DP=13;VDB=0.0074;AF1=0.4999	GT:PL:GQ	0/1:35,0,216:36
32	chromosome	660031	.	C	T	8.64	.	DP=11;VDB=0.0063;AF1=0.5;AC	GT:PL:GQ	0/1:38,0,197:40
33	chromosome	792963	.	T	C	4.13	.	DP=14;VDB=0.0153;AF1=0.4998	GT:PL:GQ	0/1:32,0,200:31
34	chromosome	887422	.	G	A	10.4	.	DP=11;VDB=0.0070;AF1=0.5;AC	GT:PL:GQ	0/1:40,0,188:42
35	chromosome	913470	.	T	C	14.2	.	DP=9;VDB=0.0056;AF1=0.5;AC1	GT:PL:GQ	0/1:44,0,144:47
36	chromosome	965171	.	T	C	17.1	.	DP=20;VDB=0.0071;AF1=0.5;AC	GT:PL:GQ	0/1:47,0,249:50
37	chromosome	1134950	.	A	G	7.8	.	DP=23;VDB=0.0076;AF1=0.5;AC	GT:PL:GQ	0/1:37,0,255:39
38	chromosome	1160048	.	A	G	29	.	DP=17;VDB=0.0054;AF1=0.5;AC	GT:PL:GQ	0/1:59,0,255:62
39	chromosome	1160095	.	A	G	19.1	.	DP=10;VDB=0.0033;AF1=0.5;AC	GT:PL:GQ	0/1:49,0,191:52
40	chromosome	1240248	.	GCCCC	GCCCCC	214	.	INDEL;DP=48;VDB=0.0137;AF1=	GT:PL:GQ	1/1:255,128,0:99
41	chromosome	1455552	.	CAAAA	CAAAAA	214	.	INDEL;DP=58;VDB=0.0164;AF1=	GT:PL:GQ	1/1:255,130,0:99
42	chromosome	1475721	.	T	C	52	.	DP=19;VDB=0.0021;AF1=0.5;AC	GT:PL:GQ	0/1:82,0,246:85
43	chromosome	1752680	.	G	A	29	.	DP=24;VDB=0.0129;AF1=0.5;AC	GT:PL:GQ	0/1:59,0,255:62
44	chromosome	1941380	.	T	C	7.8	.	DP=13;VDB=0.0046;AF1=0.5;AC	GT:PL:GQ	0/1:37,0,214:39
45	chromosome	2052955	.	A	G	12.3	.	DP=8;VDB=0.0084;AF1=0.5;AC1	GT:PL:GQ	0/1:42,0,103:44
46	chromosome	2054084	.	TC	T	13.7	.	INDEL;DP=8;VDB=0.0009;AF1=0	GT:PL:GQ	0/1:51,0,174:54
47	chromosome	2056279	.	A	G	24	.	DP=6;VDB=0.0057;AF1=0.5;AC1	GT:PL:GQ	0/1:54,0,113:57
48	chromosome	2057564	.	T	C	8.64	.	DP=10;VDB=0.0041;AF1=0.5;AC	GT:PL:GQ	0/1:38,0,129:40
49	chromosome	2058145	.	C	T,G	108	.	DP=14;VDB=0.0147;AF1=1;AC1=	GT:PL:GQ	1/1:141,39,0,141,28,138:75
50	plasmid1	37177	.	G	T	54	.	DP=49;VDB=0.0087;AF1=0.5;AC	GT:PL:GQ	0/1:84,0,255:87
51	plasmid1	66609	.	A	G	225	.	DP=55;VDB=0.0060;AF1=0.5;AC	GT:PL:GQ	0/1:255,0,255:99
52	plasmid1	66637	.	C	T	225	.	DP=71;VDB=0.0141;AF1=0.5;AC	GT:PL:GQ	0/1:255,0,249:99
53	plasmid1	66668	.	C	T	225	.	DP=80;VDB=0.0153;AF1=0.5;AC	GT:PL:GQ	0/1:255,0,255:99
54	plasmid2	2569	.	CAAAAA	CAAAA	177	.	INDEL;DP=110;VDB=0.0166;AF1	GT:PL:GQ	1/1:218,255,0:99
55	plasmid2	37805	.	TAAAAA	TAAAA	104	.	INDEL;DP=31;VDB=0.0058;AF1=	GT:PL:GQ	1/1:145,93,0:99

W18-4：全体

①は、genotype情報。0はREFのアリル、1はALTのアリルを意味する。0/1はREFとALTの両方が存在するもの(ヘテロ接合体；heterozygote)、1/1はALTのみからなるホモ接合体(homozygote)

	A	B	C	D	E	F	G	H	I	J
27	#CHROM	POS	ID	REF	ALT	QUAL	FILTER	INFO	FORMAT	①me/w3pipeline/refdata/
28	chromosome	141598	.	A	G	17.1	.	DP=7;VDB=0.0077;AF1=0.5;AC1	GT:PL:GQ	0/1:47,0,138:50
29	chromosome	242819	.	T	C	20	.	DP=9;VDB=0.0111;AF1=0.5;AC1	GT:PL:GQ	0/1:50,0,139:53
30	chromosome	359141	.	T	C	4.13	.	DP=14;VDB=0.0037;AF1=0.4998	GT:PL:GQ	0/1:32,0,187:31
31	chromosome	463882	.	T	G	6.2	.	DP=13;VDB=0.0074;AF1=0.4999	GT:PL:GQ	0/1:35,0,216:36
32	chromosome	660031	.	C	T	8.64	.	DP=11;VDB=0.0063;AF1=0.5;AC	GT:PL:GQ	0/1:38,0,197:40
33	chromosome	792963	.	T	C	4.13	.	DP=14;VDB=0.0153;AF1=0.4998	GT:PL:GQ	0/1:32,0,200:31
34	chromosome	887422	.	G	A	10.4	.	DP=11;VDB=0.0070;AF1=0.5;AC	GT:PL:GQ	0/1:40,0,188:42
35	chromosome	913470	.	T	C	14.2	.	DP=9;VDB=0.0056;AF1=0.5;AC1	GT:PL:GQ	0/1:44,0,144:47
36	chromosome	965171	.	T	C	17.1	.	DP=20;VDB=0.0071;AF1=0.5;AC	GT:PL:GQ	0/1:47,0,249:50
37	chromosome	1134950	.	A	G	7.8	.	DP=23;VDB=0.0076;AF1=0.5;AC	GT:PL:GQ	0/1:37,0,255:39
38	chromosome	1160048	.	A	G	29	.	DP=17;VDB=0.0054;AF1=0.5;AC	GT:PL:GQ	0/1:59,0,255:62
39	chromosome	1160095	.	A	G	19.1	.	DP=10;VDB=0.0033;AF1=0.5;AC	GT:PL:GQ	0/1:49,0,191:52
40	chromosome	1240248	.	GCCCC	GCCCCC	214	.	INDEL;DP=48;VDB=0.0137;AF1=	GT:PL:GQ	1/1:255,128,0:99
41	chromosome	1455552	.	CAAAA	CAAAAA	214	.	INDEL;DP=58;VDB=0.0164;AF1=	GT:PL:GQ	1/1:255,130,0:99
42	chromosome	1475721	.	T	C	52	.	DP=19;VDB=0.0021;AF1=0.5;AC	GT:PL:GQ	0/1:82,0,246:85
43	chromosome	1752680	.	G	A	29	.	DP=24;VDB=0.0129;AF1=0.5;AC	GT:PL:GQ	0/1:59,0,255:62
44	chromosome	1941380	.	T	C	7.8	.	DP=13;VDB=0.0046;AF1=0.5;AC	GT:PL:GQ	0/1:37,0,214:39
45	chromosome	2052955	.	A	G	12.3	.	DP=8;VDB=0.0084;AF1=0.5;AC1	GT:PL:GQ	0/1:42,0,103:44
46	chromosome	2054084	.	TC	T	13.7	.	INDEL;DP=8;VDB=0.0009;AF1=0	GT:PL:GQ	0/1:51,0,174:54
47	chromosome	2056279	.	A	G	24	.	DP=6;VDB=0.0057;AF1=0.5;AC1	GT:PL:GQ	0/1:54,0,113:57
48	chromosome	2057564	.	T	C	8.64	.	DP=10;VDB=0.0041;AF1=0.5;AC	GT:PL:GQ	0/1:38,0,129:40
49	chromosome	2058145	.	C	T,G	108	.	DP=14;VDB=0.0147;AF1=1;AC1=	GT:PL:GQ	1/1:141,39,0,141,28,138:75
50	plasmid1	37177	.	G	T	54	.	DP=49;VDB=0.0087;AF1=0.5;AC	GT:PL:GQ	0/1:84,0,255:87
51	plasmid1	66609	.	A	G	225	.	DP=55;VDB=0.0060;AF1=0.5;AC	GT:PL:GQ	0/1:255,0,255:99
52	plasmid1	66637	.	C	T	225	.	DP=71;VDB=0.0141;AF1=0.5;AC	GT:PL:GQ	0/1:255,0,249:99
53	plasmid1	66668	.	C	T	225	.	DP=80;VDB=0.0153;AF1=0.5;AC	GT:PL:GQ	0/1:255,0,255:99
54	plasmid2	2569	.	CAAAAA	CAAAA	177	.	INDEL;DP=110;VDB=0.0166;AF1	GT:PL:GQ	1/1:218,255,0:99
55	plasmid2	37805	.	TAAAAA	TAAAA	104	.	INDEL;DP=31;VDB=0.0058;AF1=	GT:PL:GQ	1/1:145,93,0:99

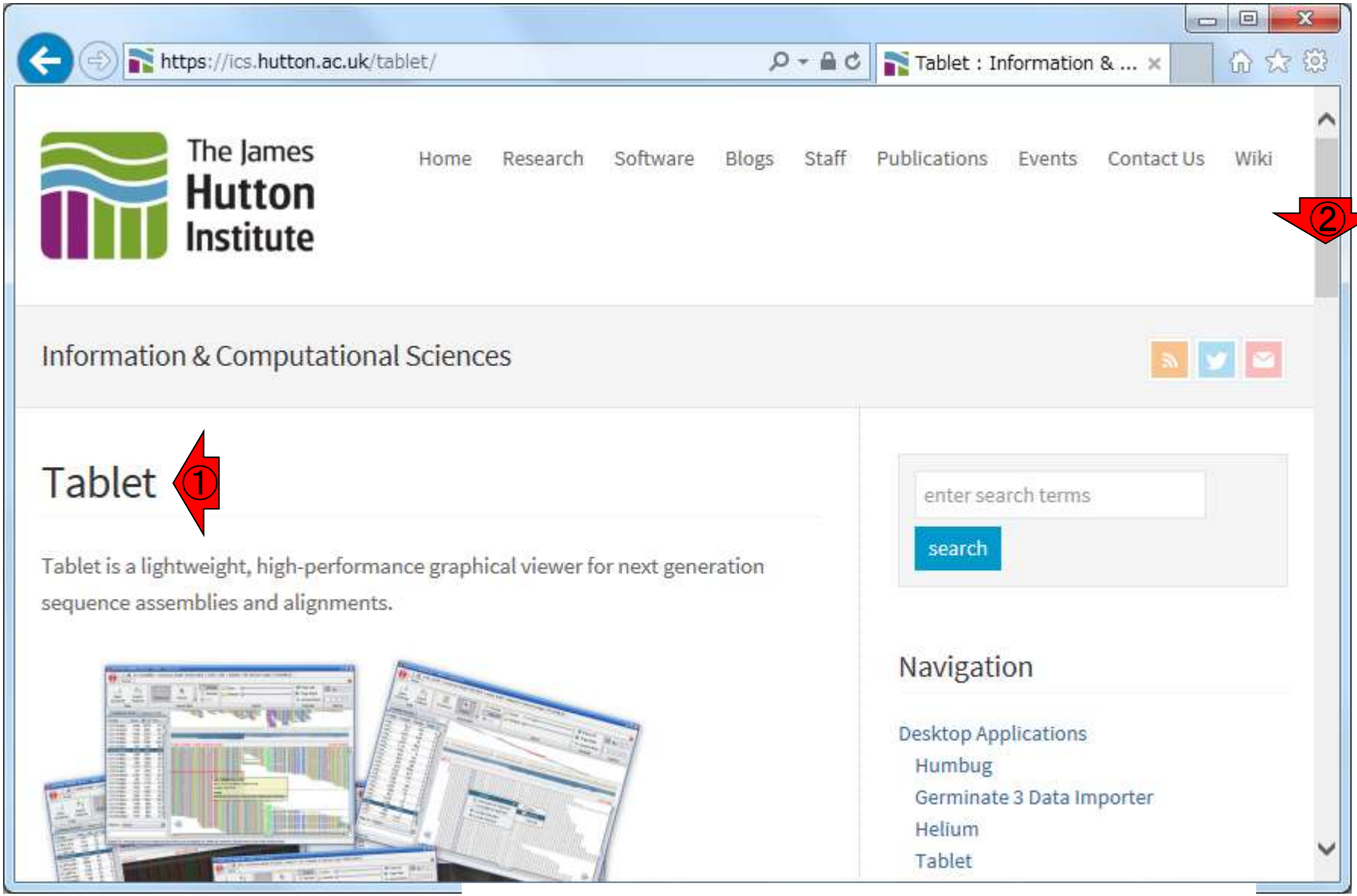
W18-4：全体

我々の目的である概要配列の修正箇所の有力候補は赤矢印部分だということ!

	A	B	C	D	E	F	G	H	I	J
27	#CHROM	POS	ID	REF	ALT	QUAL	FILTER	INFO	FORMAT	/home/w3pipeline/refdata/
28	chromosome	141598	.	A	G	17.1	.	DP=7;VDB=0.0077;AF1=0.5;AC1	GT:PL:GQ	0/1:47,0,138:50
29	chromosome	242819	.	T	C	20	.	DP=9;VDB=0.0111;AF1=0.5;AC1	GT:PL:GQ	0/1:50,0,139:53
30	chromosome	359141	.	T	C	4.13	.	DP=14;VDB=0.0037;AF1=0.4998	GT:PL:GQ	0/1:32,0,187:31
31	chromosome	463882	.	T	G	6.2	.	DP=13;VDB=0.0074;AF1=0.4999	GT:PL:GQ	0/1:35,0,216:36
32	chromosome	660031	.	C	T	8.64	.	DP=11;VDB=0.0063;AF1=0.5;AC	GT:PL:GQ	0/1:38,0,197:40
33	chromosome	792963	.	T	C	4.13	.	DP=14;VDB=0.0153;AF1=0.4998	GT:PL:GQ	0/1:32,0,200:31
34	chromosome	887422	.	G	A	10.4	.	DP=11;VDB=0.0070;AF1=0.5;AC	GT:PL:GQ	0/1:40,0,188:42
35	chromosome	913470	.	T	C	14.2	.	DP=9;VDB=0.0056;AF1=0.5;AC1	GT:PL:GQ	0/1:44,0,144:47
36	chromosome	965171	.	T	C	17.1	.	DP=20;VDB=0.0071;AF1=0.5;AC	GT:PL:GQ	0/1:47,0,249:50
37	chromosome	1134950	.	A	G	7.8	.	DP=23;VDB=0.0076;AF1=0.5;AC	GT:PL:GQ	0/1:37,0,255:39
38	chromosome	1160048	.	A	G	29	.	DP=17;VDB=0.0054;AF1=0.5;AC	GT:PL:GQ	0/1:59,0,255:62
39	chromosome	1160095	.	A	G	19.1	.	DP=10;VDB=0.0033;AF1=0.5;AC	GT:PL:GQ	0/1:49,0,191:52
40	chromosome	1240248	.	GCCCC	GCCCCC	214	.	INDEL;DP=48;VDB=0.0137;AF1=	GT:PL:GQ	1/1:255,128,0:99
41	chromosome	1455552	.	CAAAA	CAAAAA	214	.	INDEL;DP=58;VDB=0.0164;AF1=	GT:PL:GQ	1/1:255,130,0:99
42	chromosome	1475721	.	T	C	52	.	DP=19;VDB=0.0021;AF1=0.5;AC	GT:PL:GQ	0/1:82,0,246:85
43	chromosome	1752680	.	G	A	29	.	DP=24;VDB=0.0129;AF1=0.5;AC	GT:PL:GQ	0/1:59,0,255:62
44	chromosome	1941380	.	T	C	7.8	.	DP=13;VDB=0.0046;AF1=0.5;AC	GT:PL:GQ	0/1:37,0,214:39
45	chromosome	2052955	.	A	G	12.3	.	DP=8;VDB=0.0084;AF1=0.5;AC1	GT:PL:GQ	0/1:42,0,103:44
46	chromosome	2054084	.	TC	T	13.7	.	INDEL;DP=8;VDB=0.0009;AF1=0	GT:PL:GQ	0/1:51,0,174:54
47	chromosome	2056279	.	A	G	24	.	DP=6;VDB=0.0057;AF1=0.5;AC1	GT:PL:GQ	0/1:54,0,113:57
48	chromosome	2057564	.	T	C	8.64	.	DP=10;VDB=0.0041;AF1=0.5;AC	GT:PL:GQ	0/1:38,0,129:40
49	chromosome	2058145	.	C	T,G	108	.	DP=14;VDB=0.0147;AF1=1;AC1=	GT:PL:GQ	1/1:41,39,0,141,28,138:75
50	plasmid1	37177	.	G	T	54	.	DP=49;VDB=0.0087;AF1=0.5;AC	GT:PL:GQ	0/1:84,0,255:87
51	plasmid1	66609	.	A	G	225	.	DP=55;VDB=0.0060;AF1=0.5;AC	GT:PL:GQ	0/1:255,0,255:99
52	plasmid1	66637	.	C	T	225	.	DP=71;VDB=0.0141;AF1=0.5;AC	GT:PL:GQ	0/1:255,0,249:99
53	plasmid1	66668	.	C	T	225	.	DP=80;VDB=0.0153;AF1=0.5;AC	GT:PL:GQ	0/1:255,0,255:99
54	plasmid2	2569	.	CAAAAA	CAAAA	177	.	INDEL;DP=110;VDB=0.0166;AF1	GT:PL:GQ	1/1:218,255,0:99
55	plasmid2	37805	.	TAAAAA	TAAAA	104	.	INDEL;DP=31;VDB=0.0058;AF1=	GT:PL:GQ	1/1:45,93,0:99

①Tabletのウェブページ。②ページ下部に移動

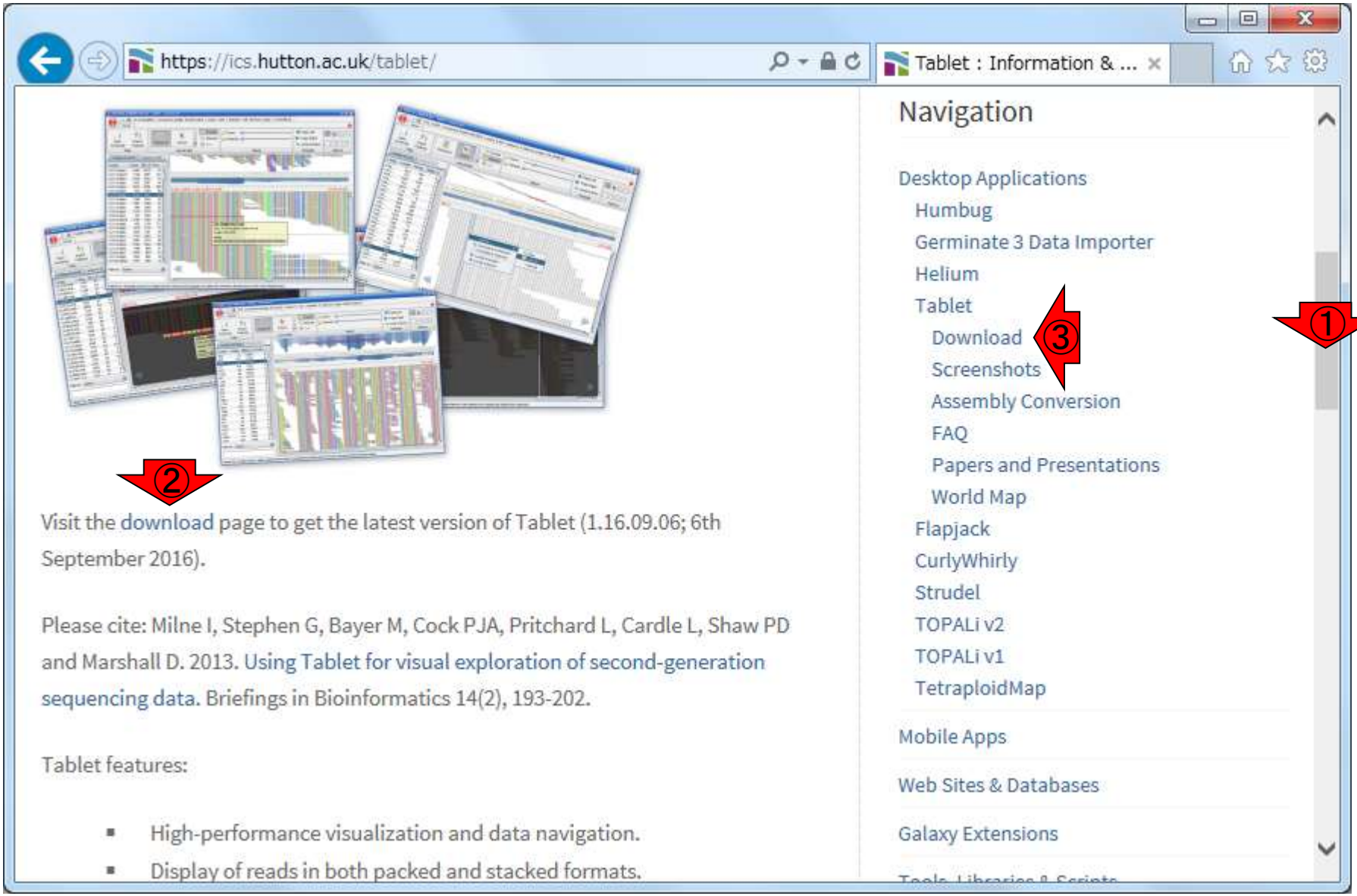
W19-1：Tablet



Milne et al., *Brief Bioinform.*, **14**: 193-202, 2013

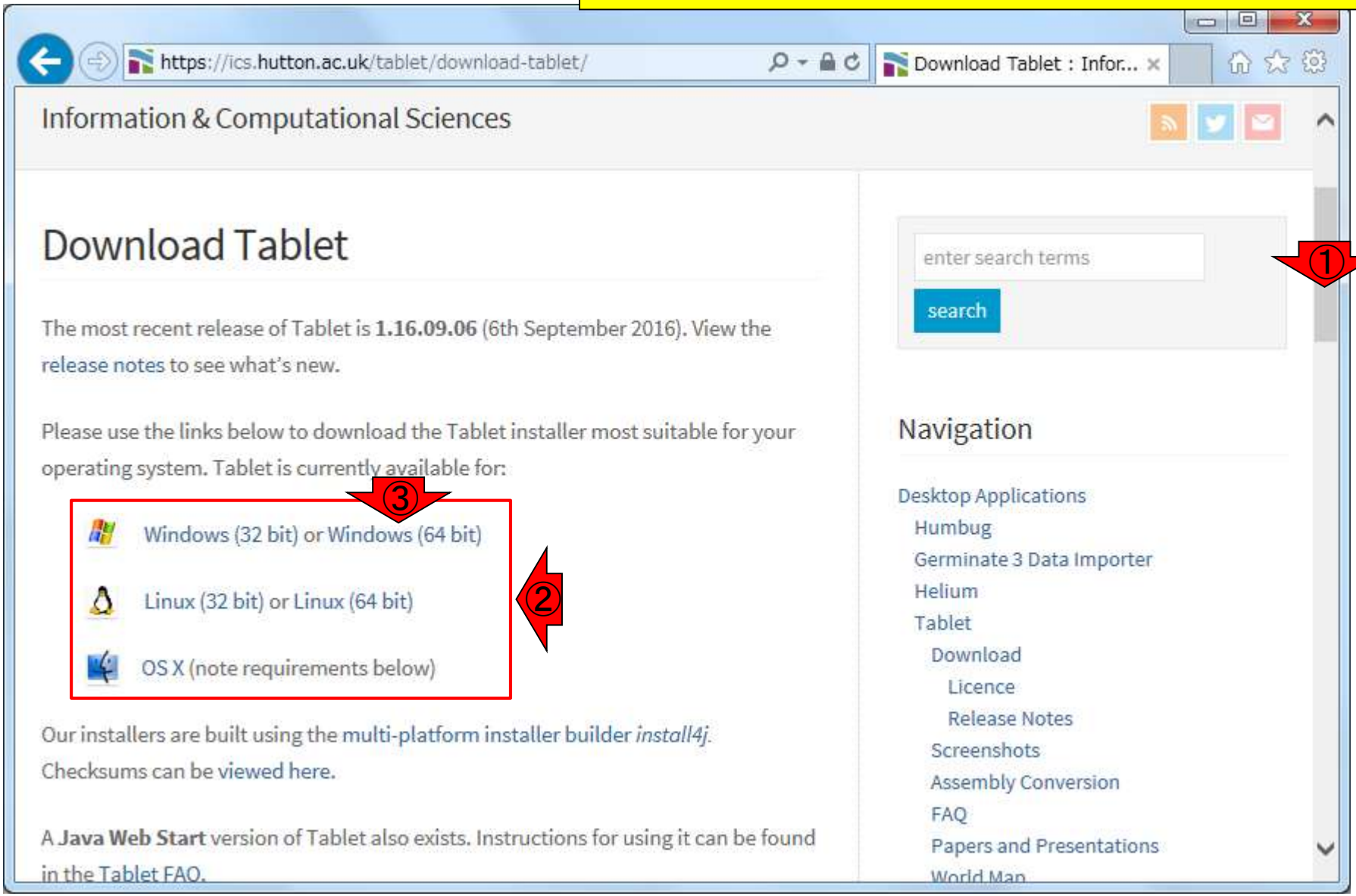
W19-1：Tablet

①このあたり。②または③を押してダウンロードページへ



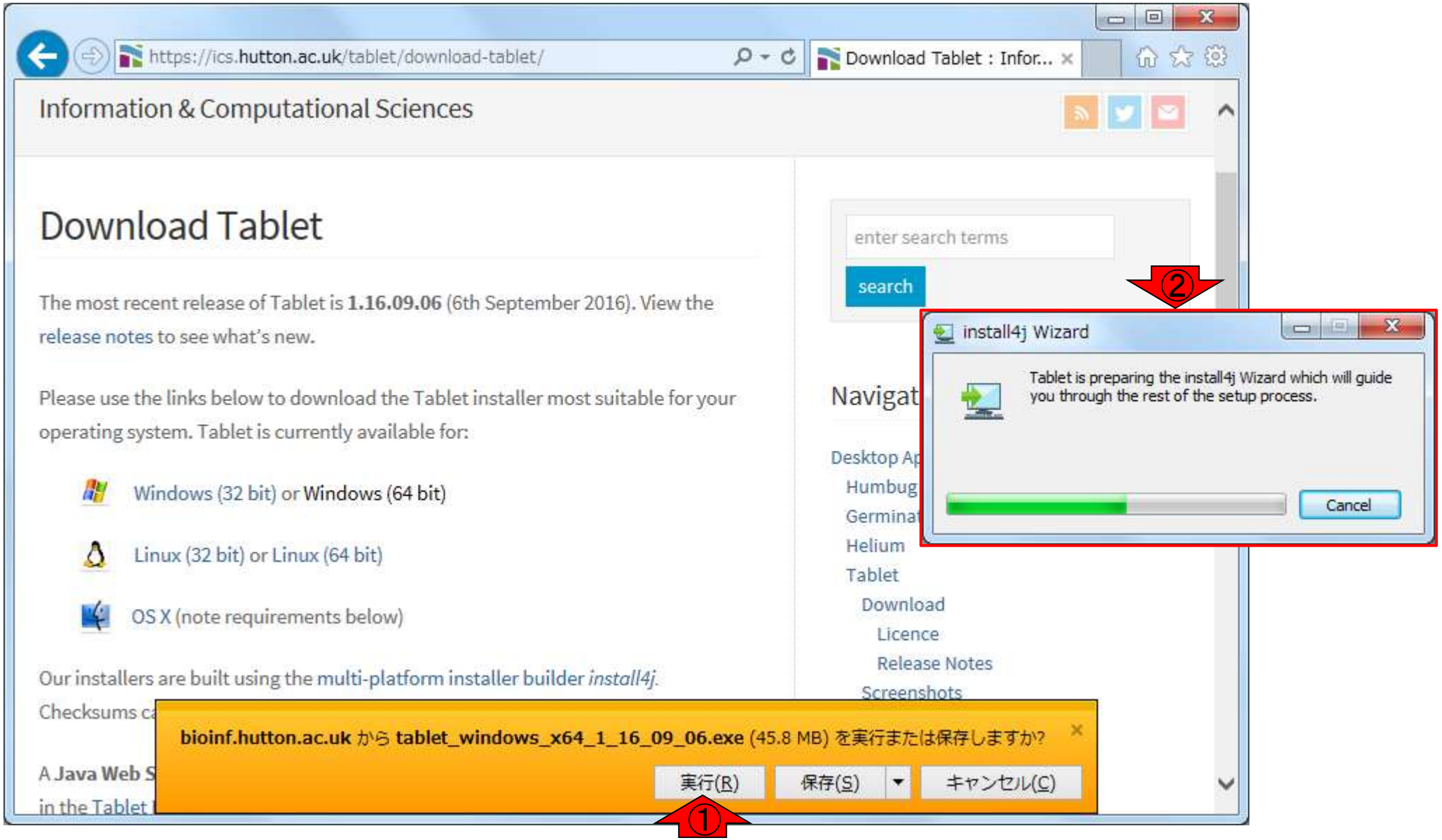
W19-1：Tablet

ダウンロードページ移動後に、①ちょっとページ下部に移動したところ。②赤枠で示すようにWindows、Linux、Macintoshのどれでも OK。ここでは③Windows (64 bit)版のインストール例を示す



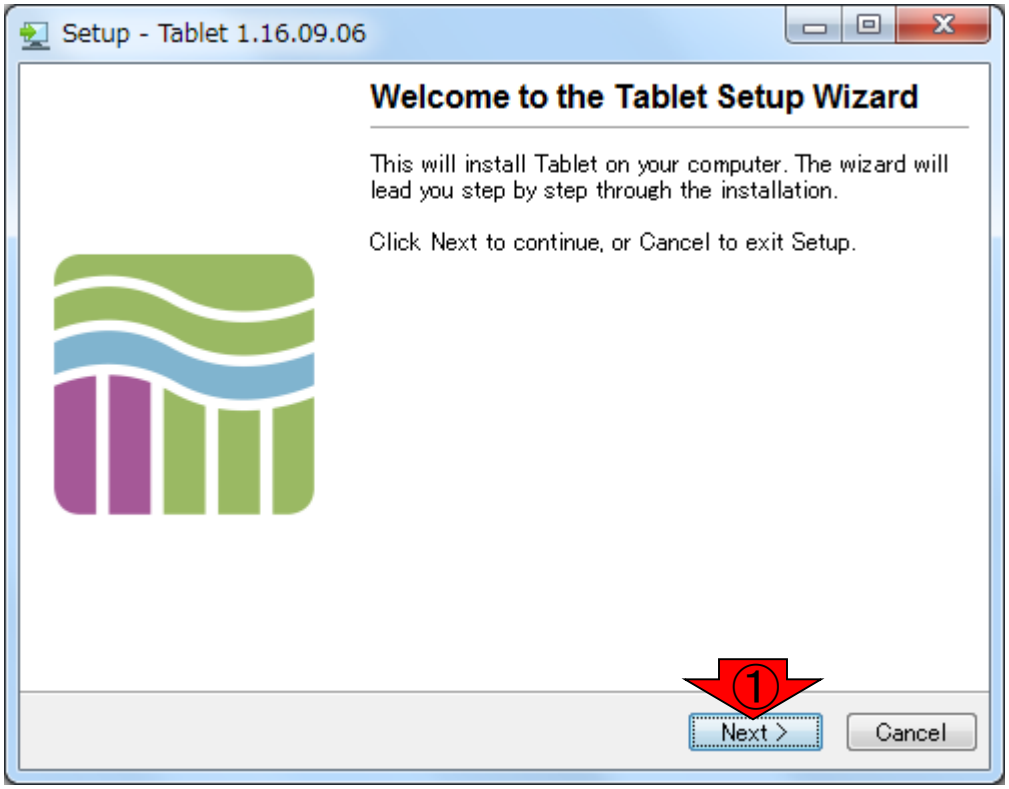
①実行。②こんなのが出ます

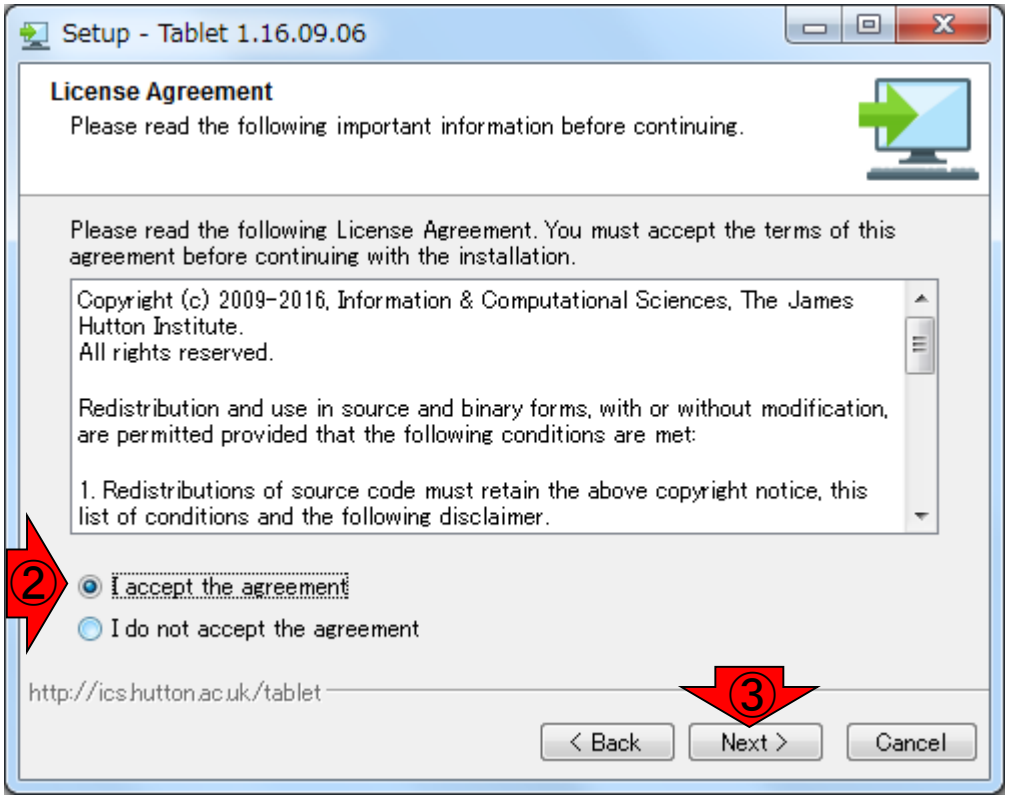
W19-1：Tablet



①Next、②I accept…、③Next

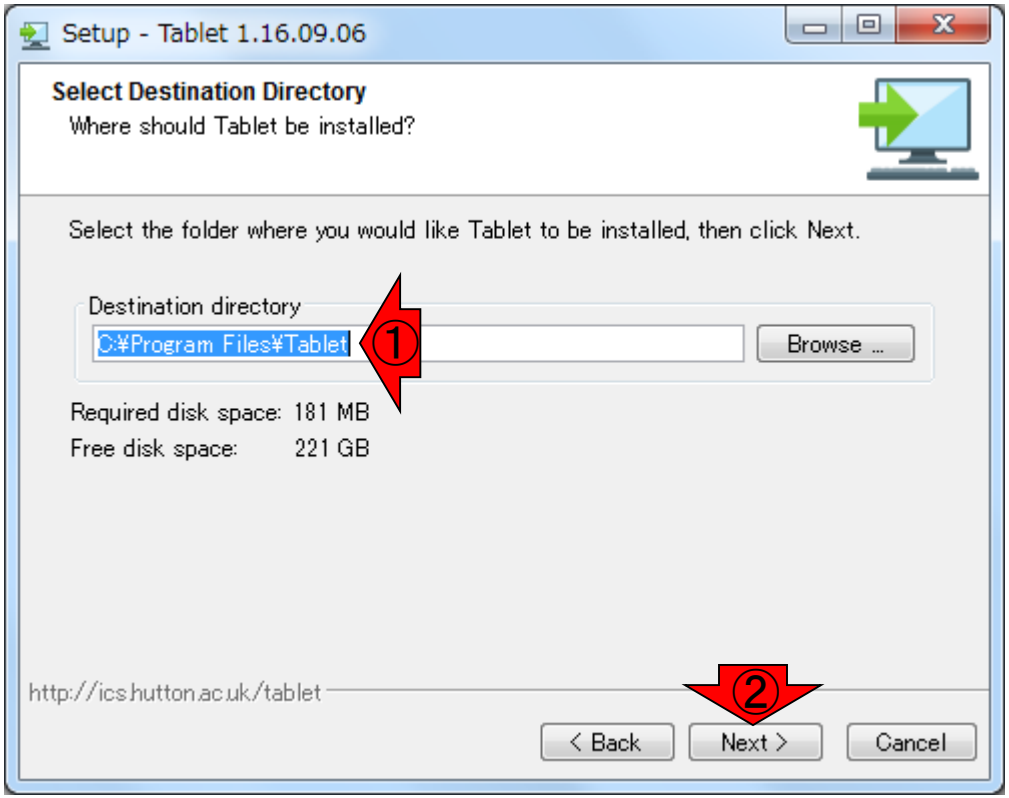
W19-1：Tablet

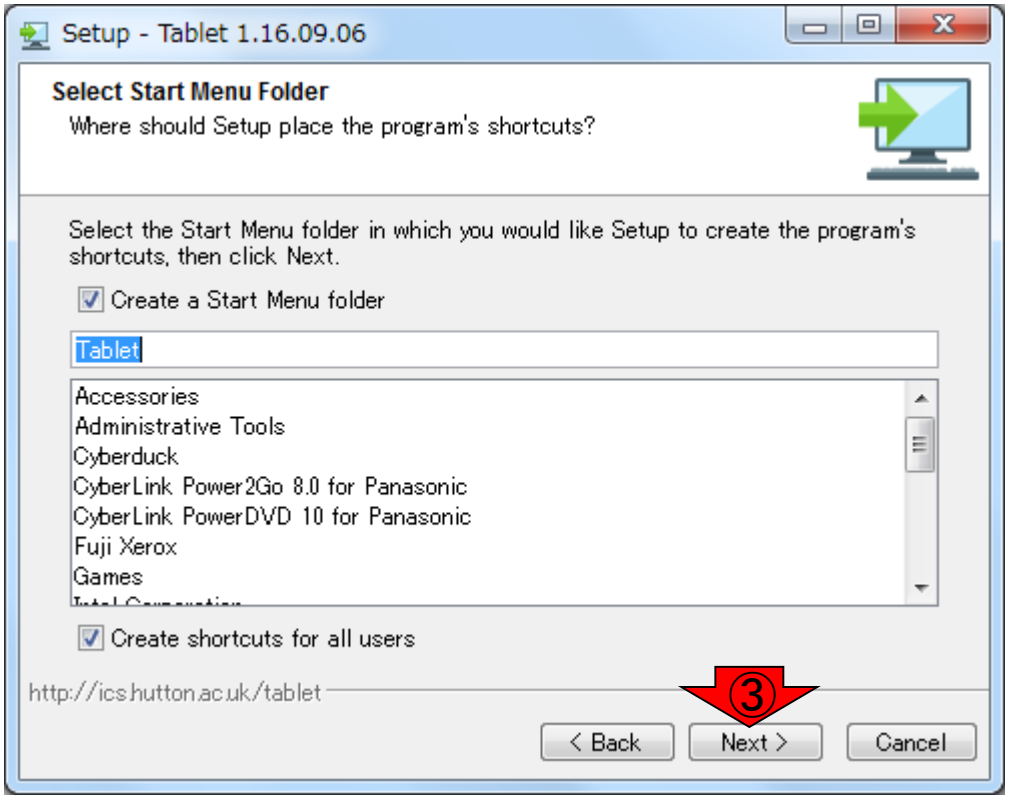




①ここにインストールされます。②Next。③Next

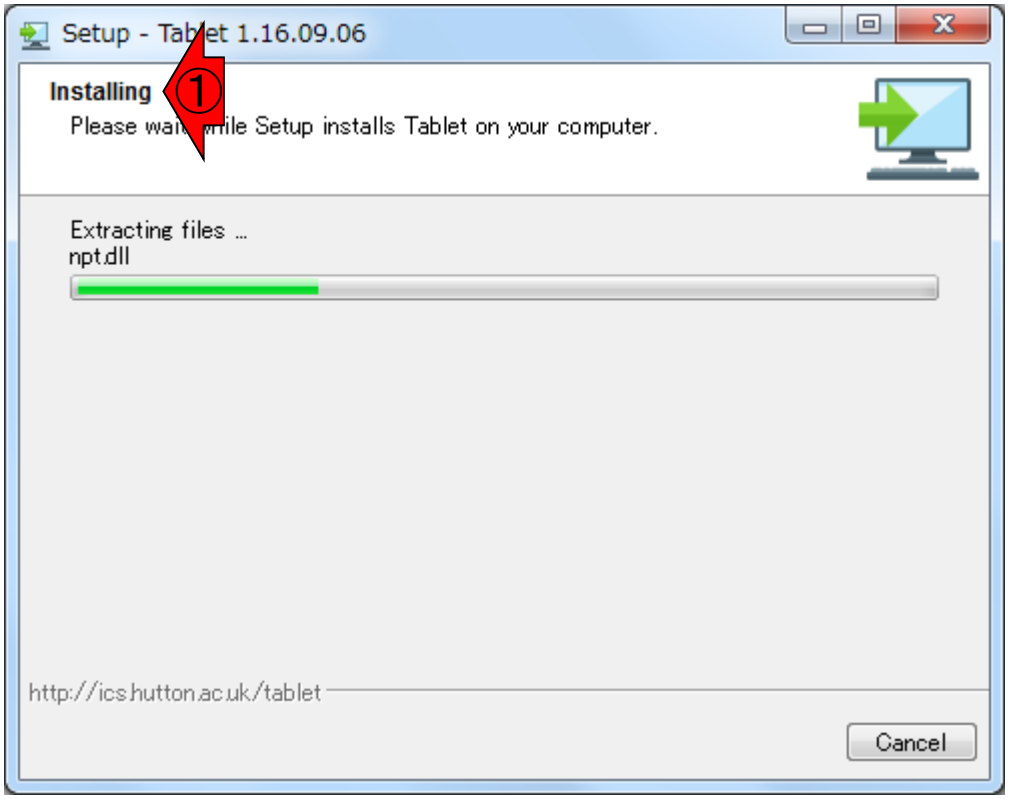
W19-1：Tablet

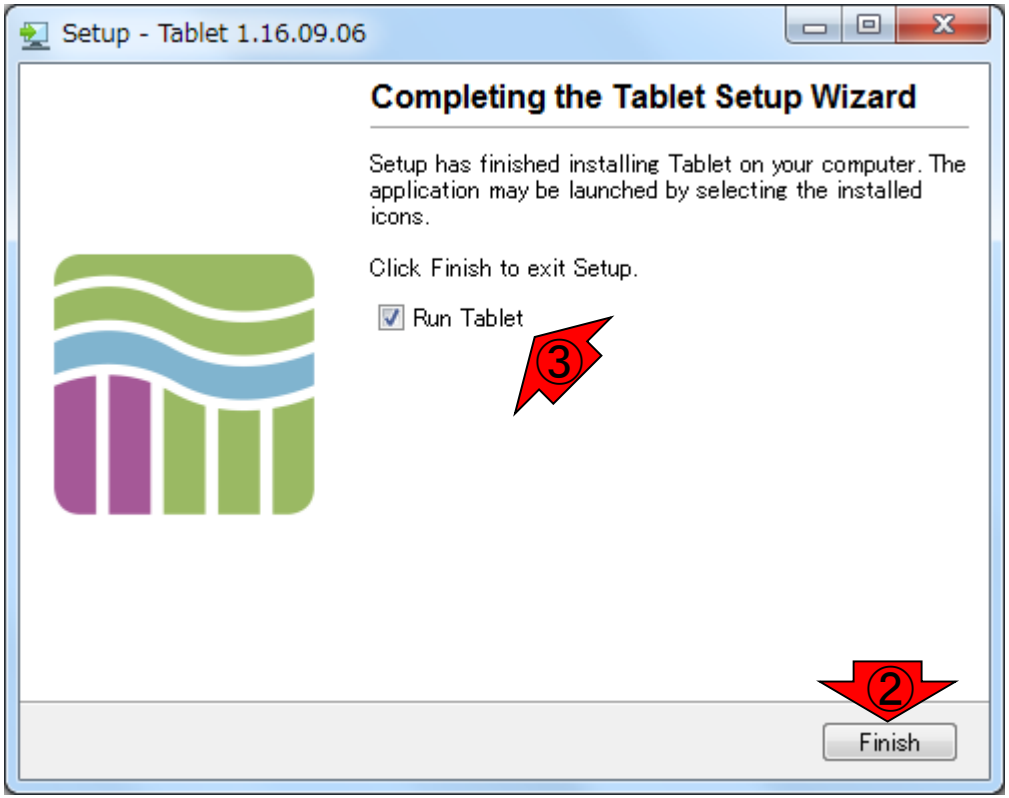




W19-1：Tablet

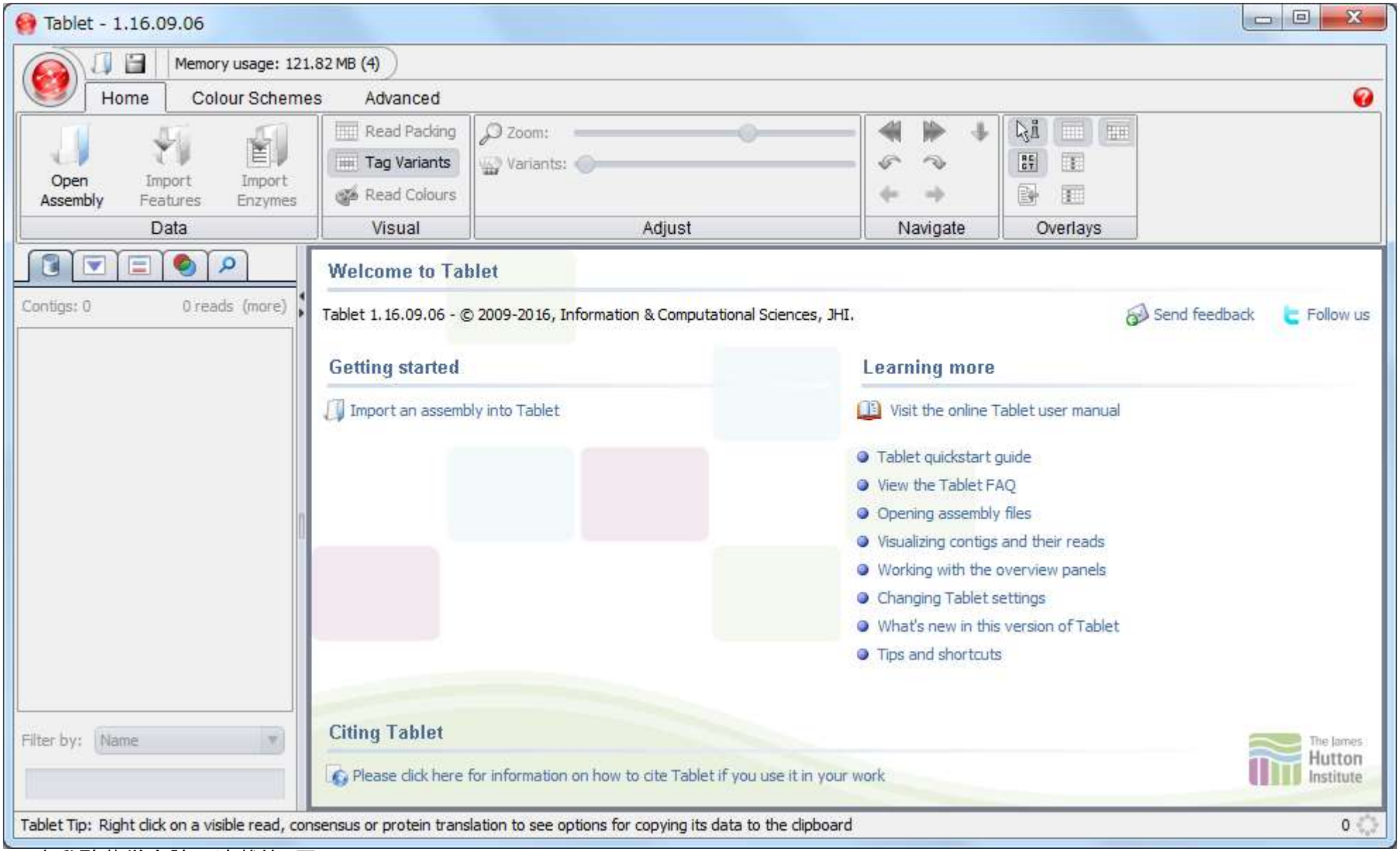
①インストール中。無事インストール完了したので、②Finish。③Run Tabletにチェックが入っているので自動的に起動する





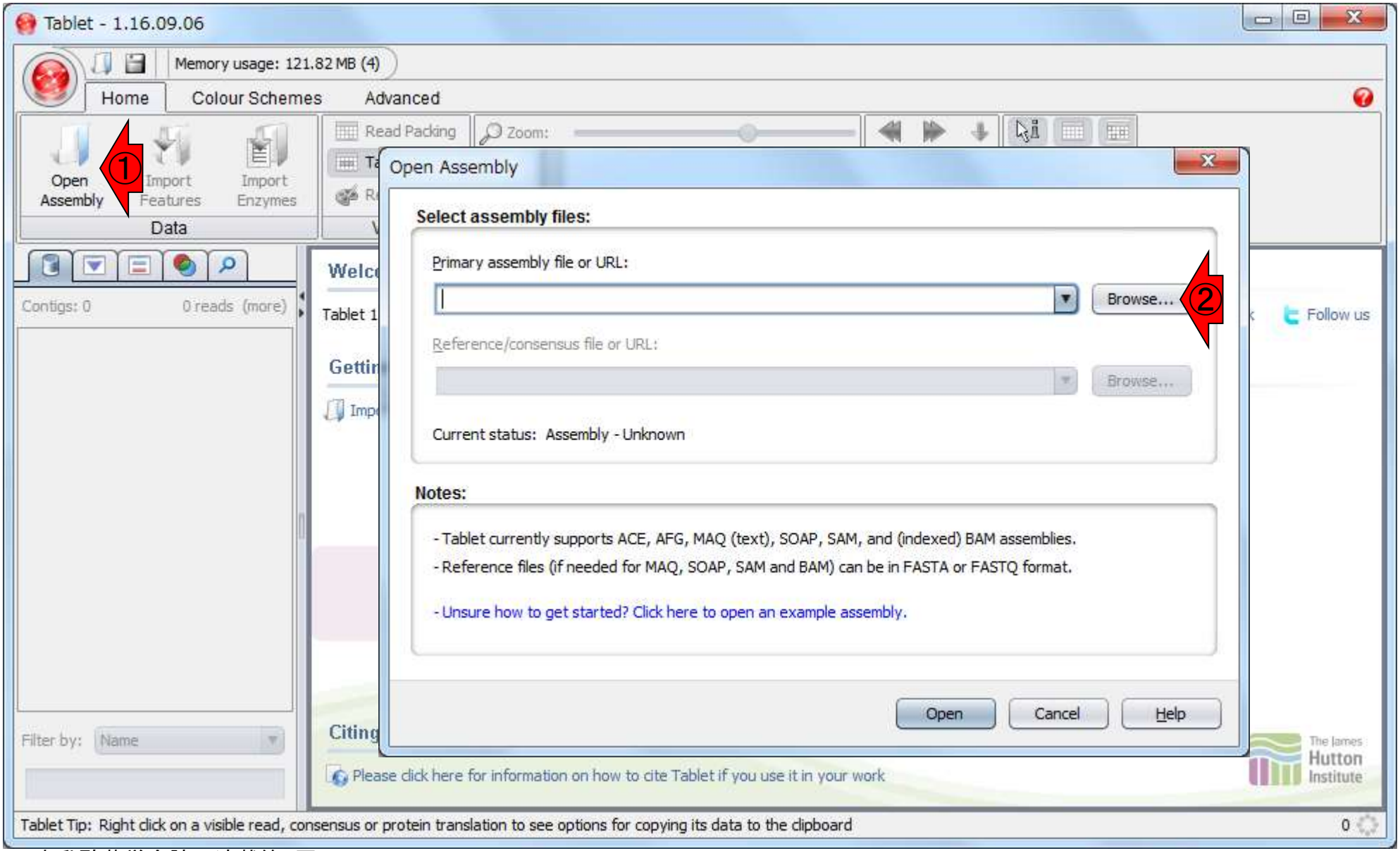
Tablet起動直後の状態はこんな感じです

W19-2：起動直後



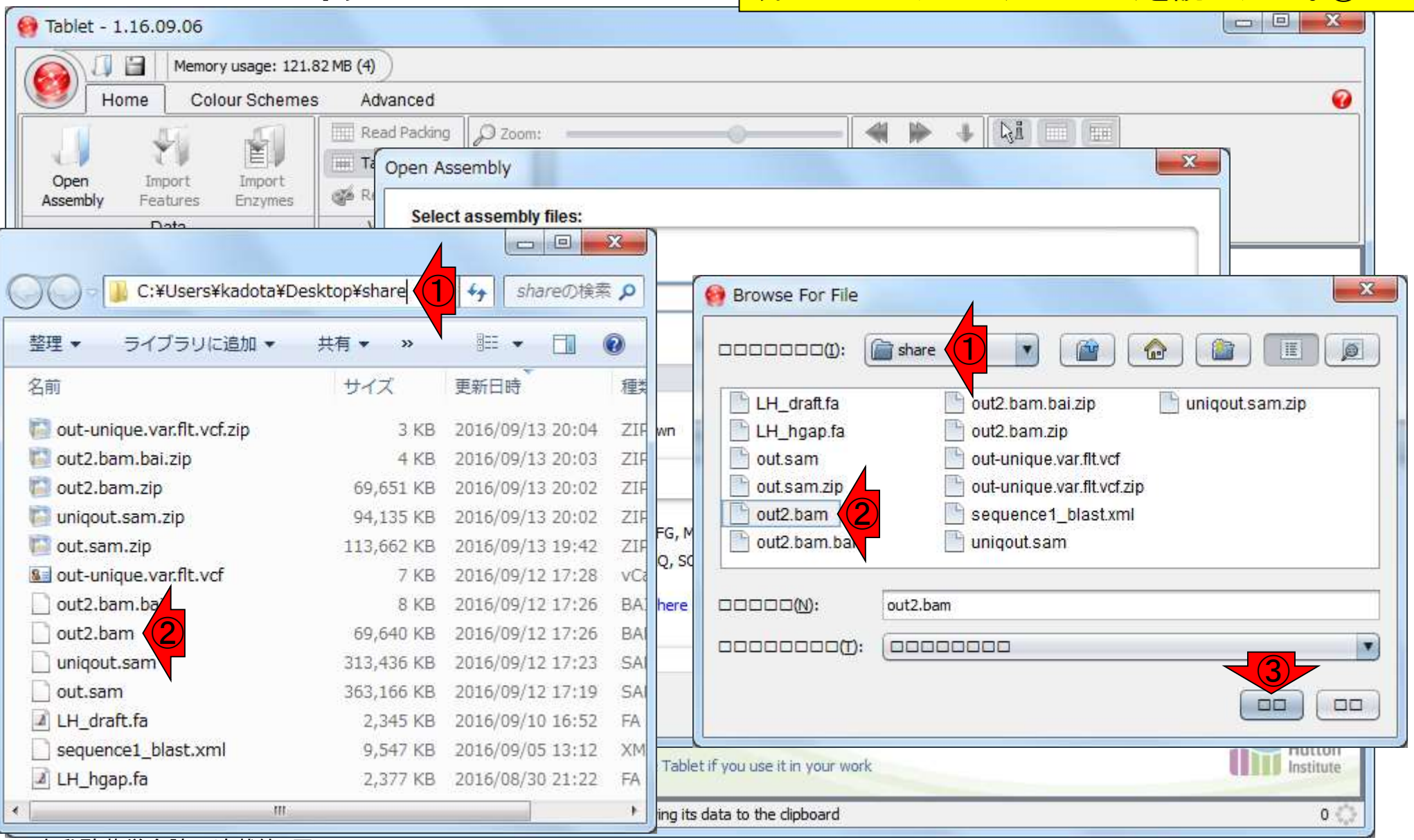
W19-3：読み込み

①Open Assemblyボタンを押して、②Browse。ここではソート済みBAMファイル(out2.bam)を読み込む



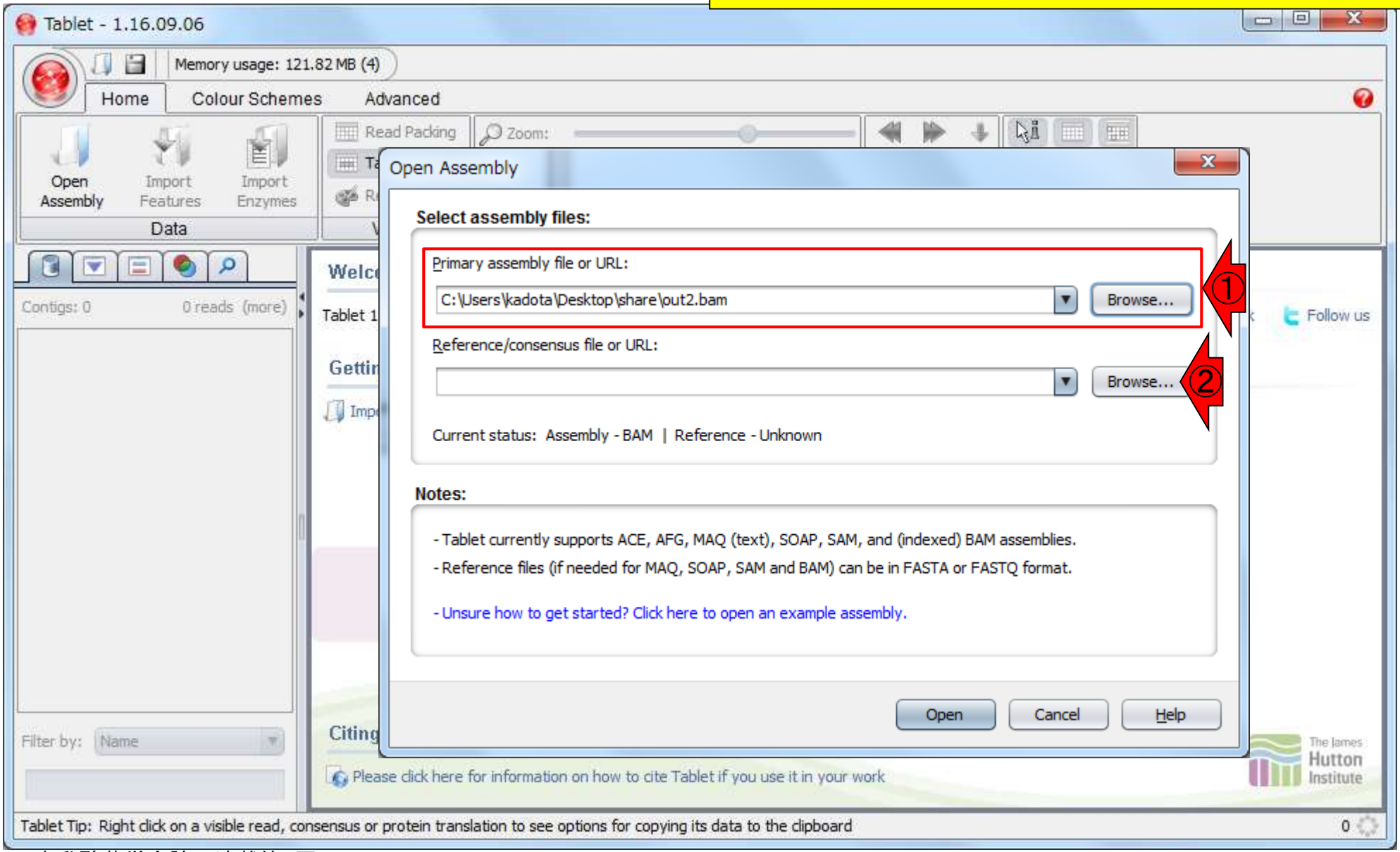
W19-3：読み込み

文字化けしているようで見づらいが、どうにかして①（ここでは）共有フォルダ上にある、②ソート済みBAMファイル(out2.bam)を読み込む。③ここ



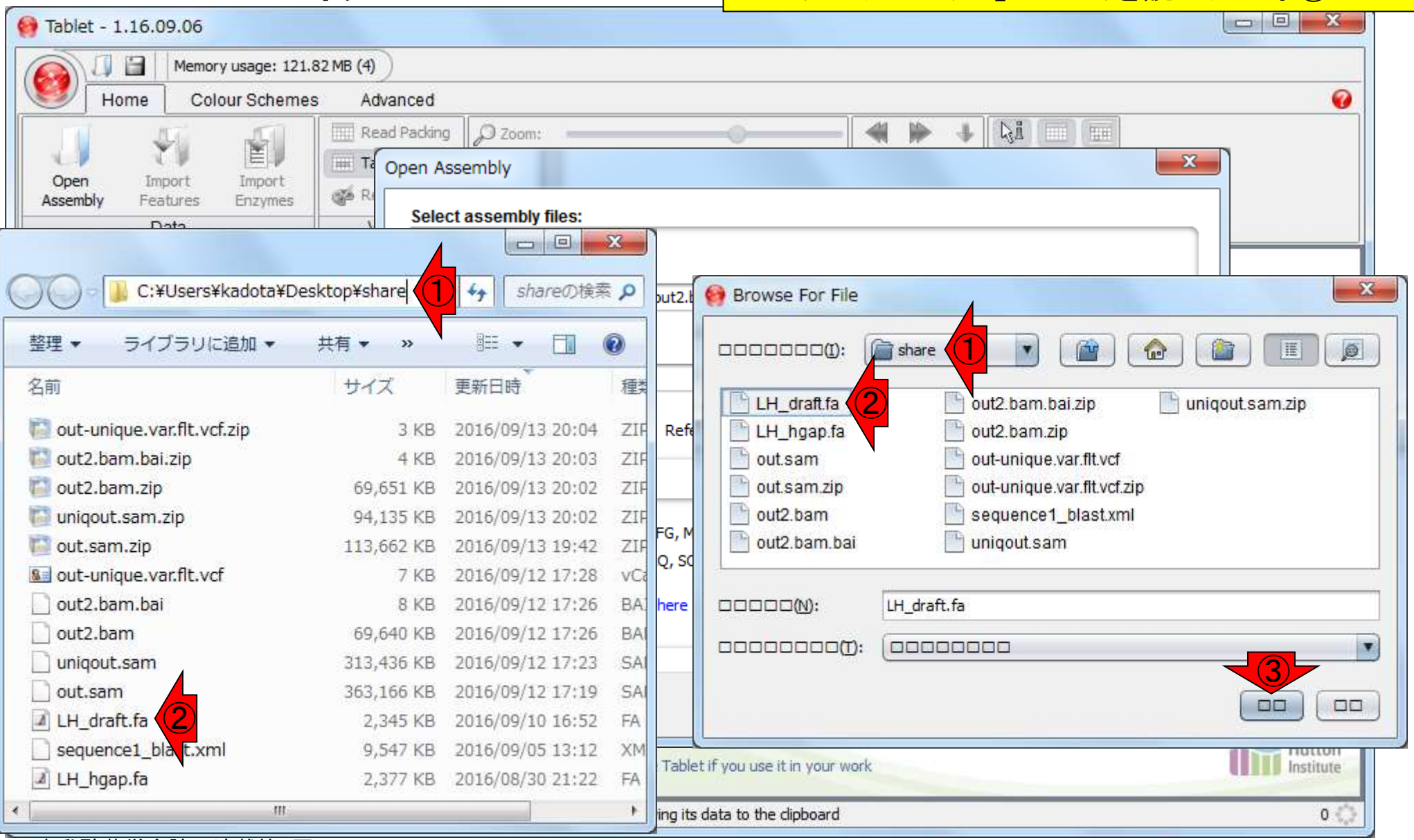
W19-3：読み込み

①ソート済みBAMファイル(out2.bam)の読み込みは成功。次はマッピングに用いたリファレンス配列ファイル(LH_draft.fa)の読み込みをすべく、②Browse



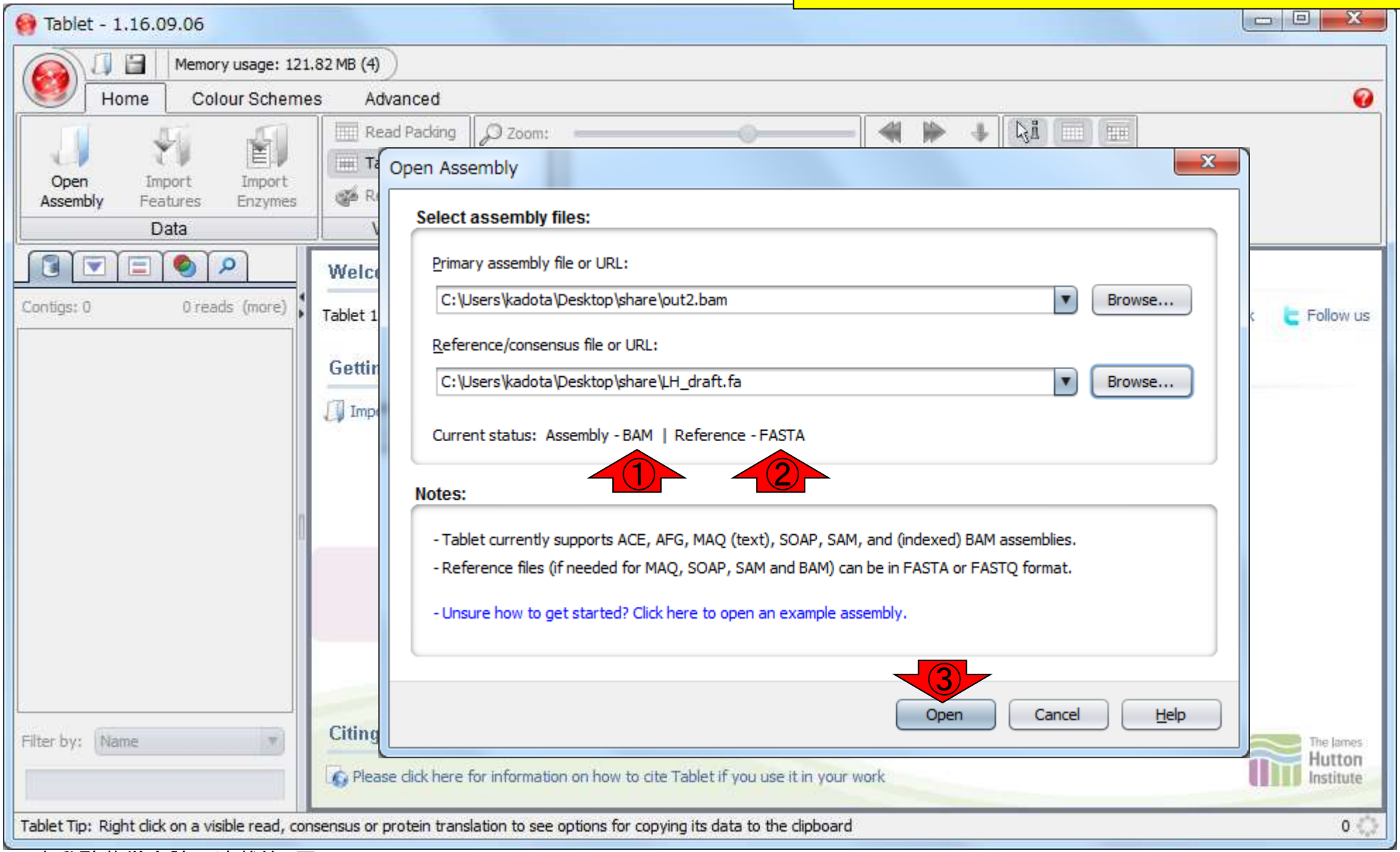
W19-3：読み込み

文字化けしているようで見づらいが、どうにかして①（ここでは）共有フォルダ上にある、②リファレンス配列ファイル(LH_draft.fa)を読み込む。③ここ



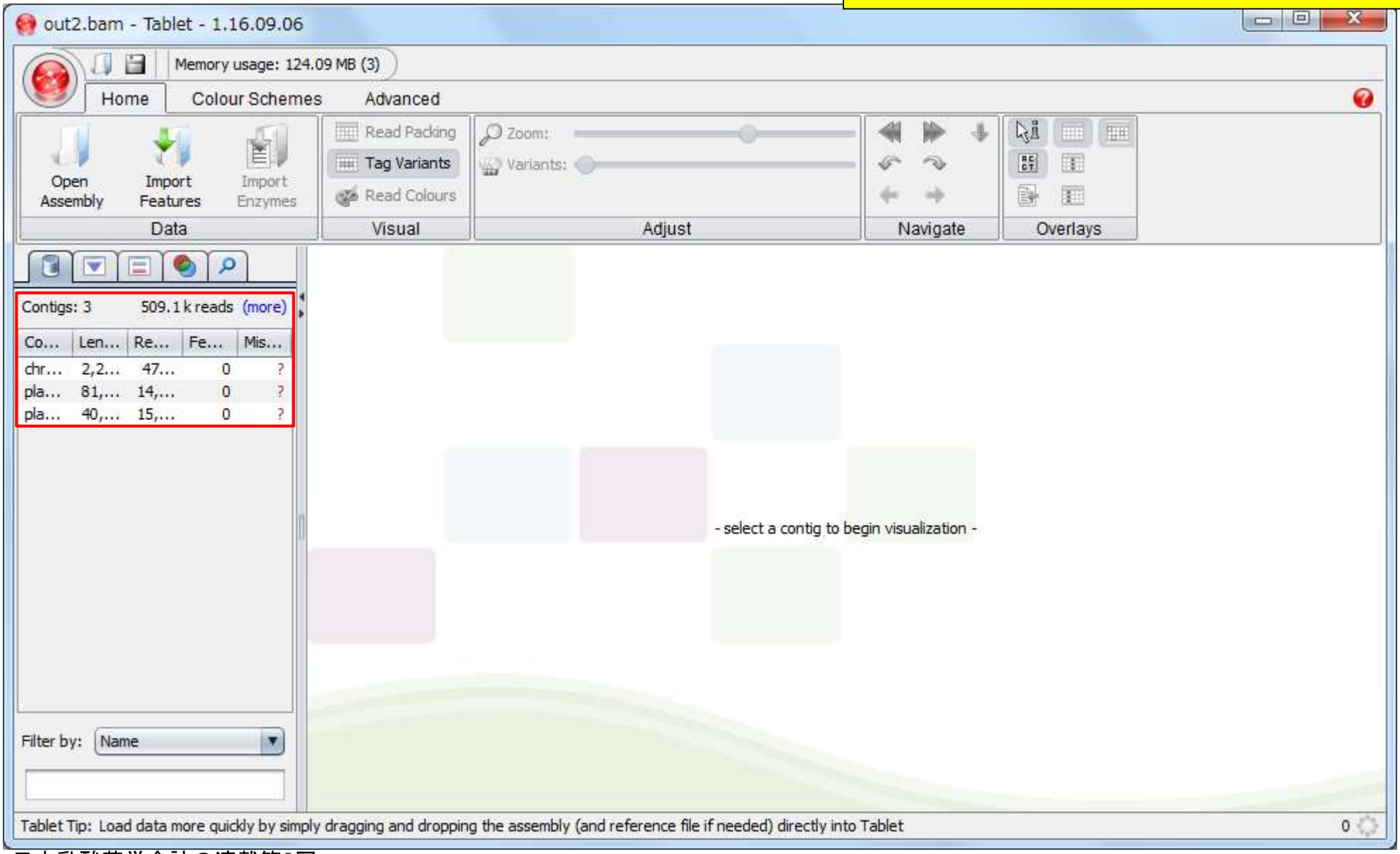
W19-3：読み込み

2つのファイルともに無事読み込み成功。よく見ると、①と②のところもunknownだったのが、BAMとFASTAに切り替わっている。③Open



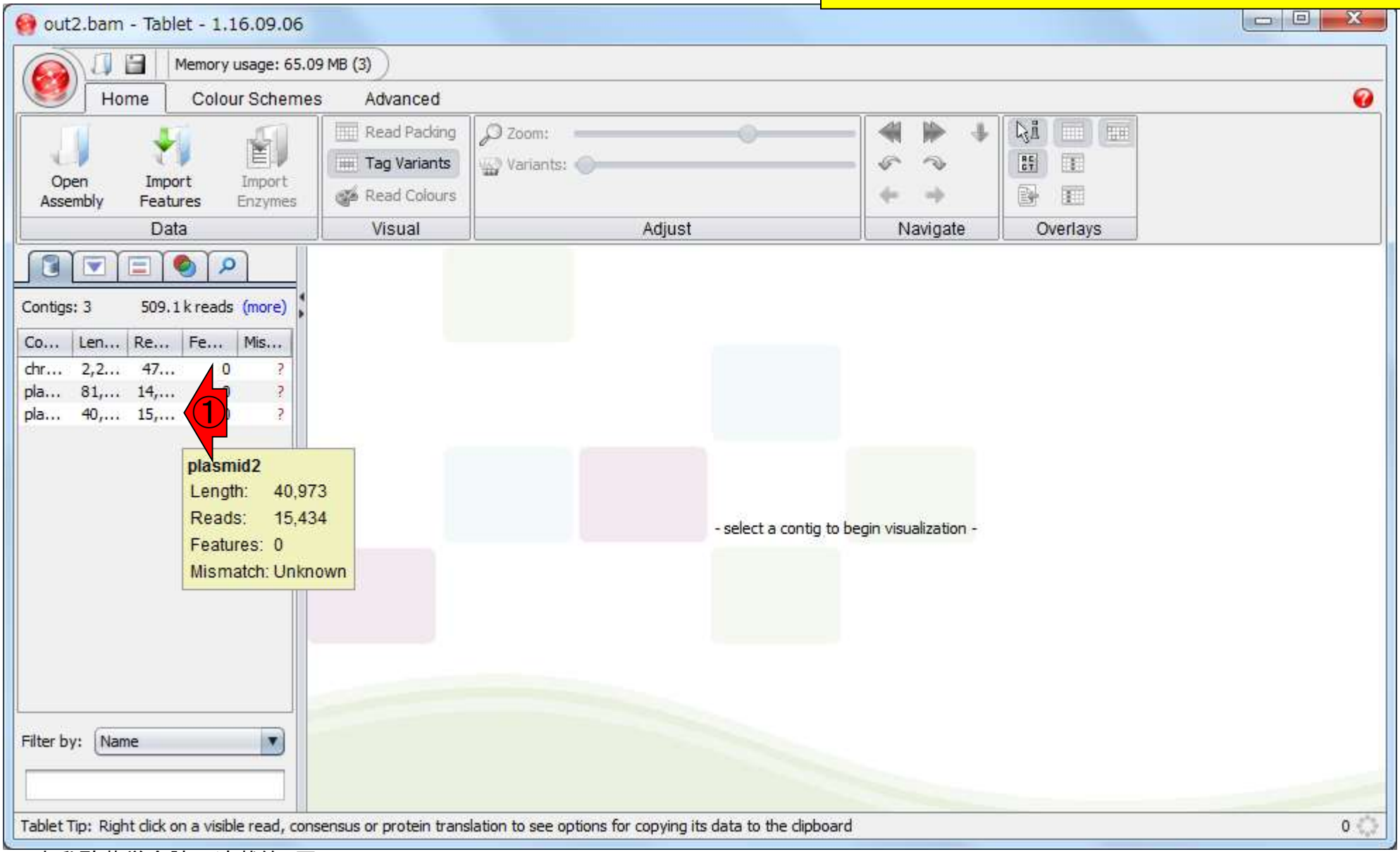
W19-3：読み込み

読み込み直後の状態。赤枠あたりが変わった。ここで見えているのは概要配列中の3つ(chromosome, plasmid1, plasmid2)



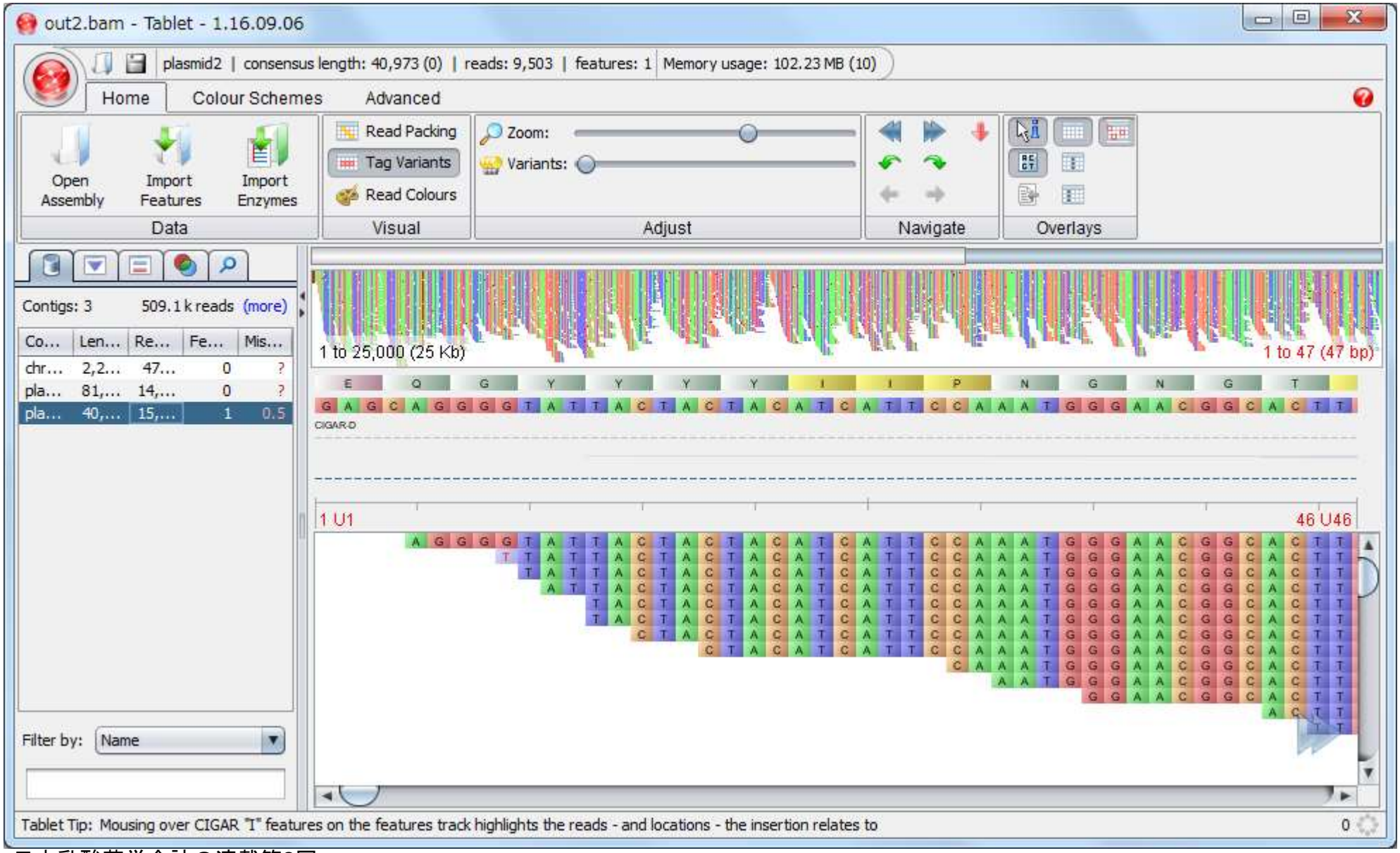
W19-4：概観

①のあたりにマウス移動させると、plasmid2の配列長(40,973 bp)とマップされたリード数(15,434)が表示される。クリックすると…



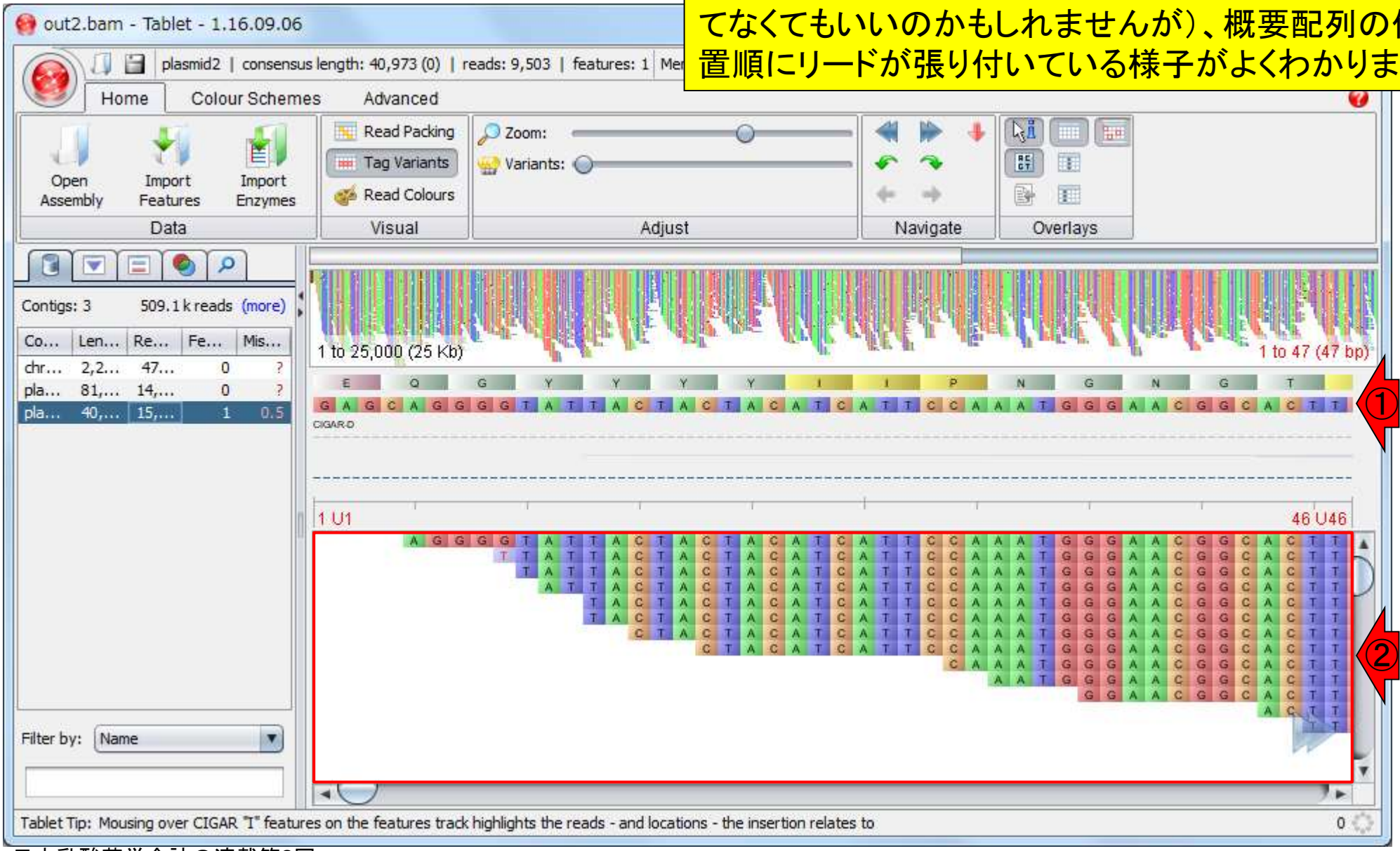
こんな感じになります

W19-4：概観



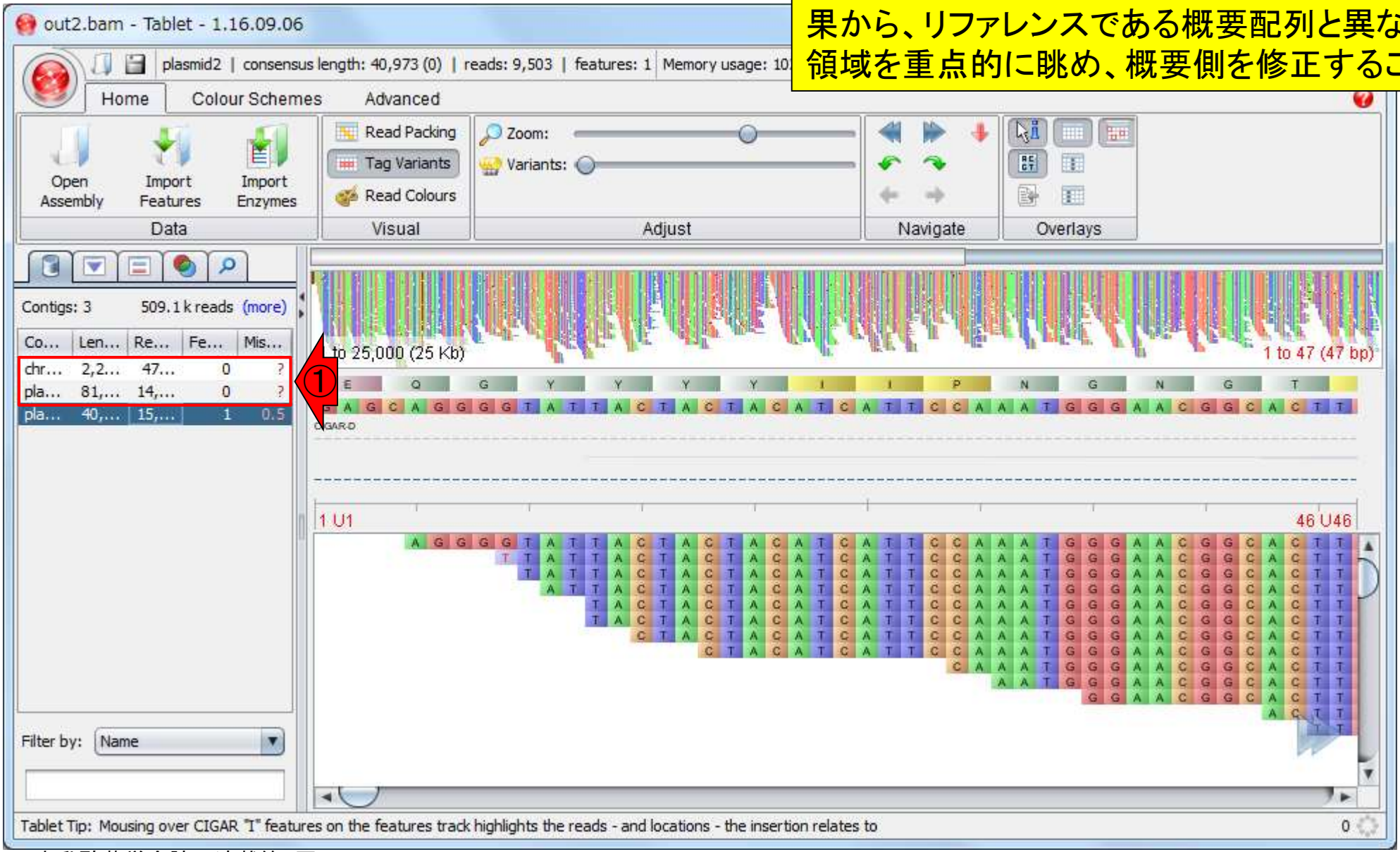
W19-4：概観

①の部分が概要配列(LH_draft.fa)。②の部分がマップされたリードの情報です。ソート済みBAMファイル(out2.bam)を読み込ませているので(本当はソートしてなくてもいいのかもしれませんが)、概要配列の位置順にリードが張り付いている様子がよくわかります



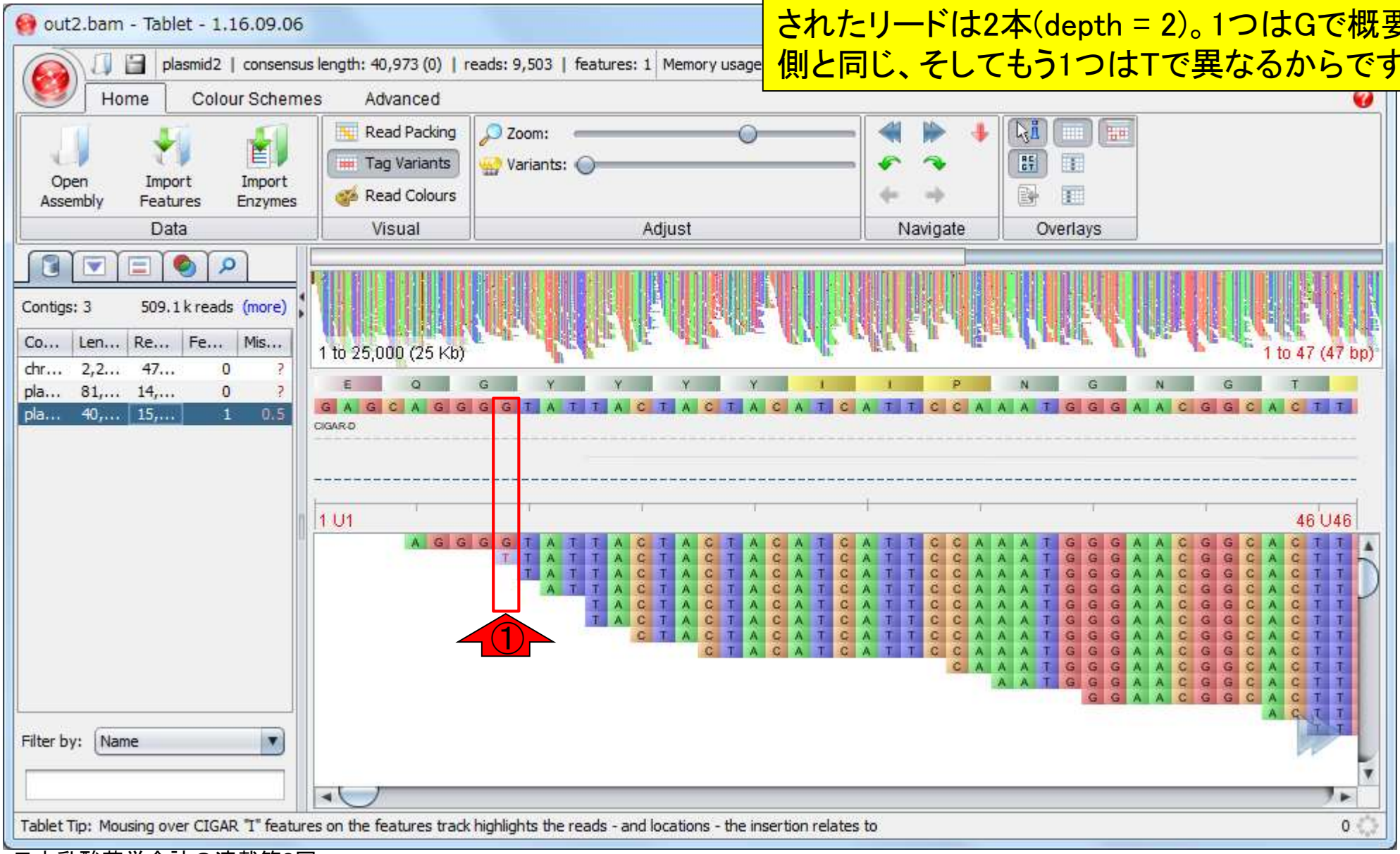
W19-5：目的おさらい

他の配列（chromosomeとplasmid1）についても同様。我々が今やりたいのは、ミスマッチやindel（挿入や欠失）を許容してマップした結果から、リファレンスである概要配列と異なる領域を重点的に眺め、概要側を修正すること



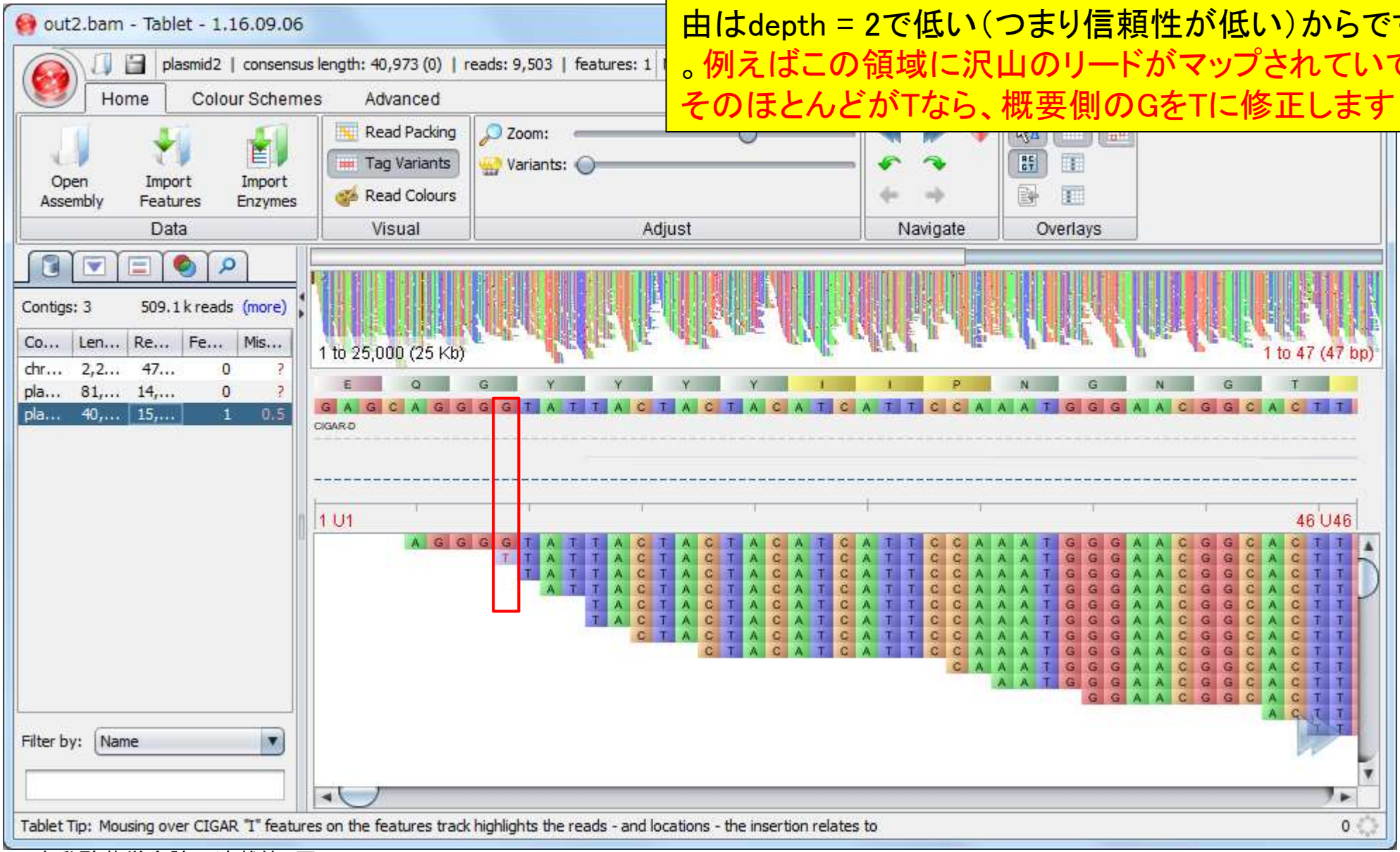
W19-5: 目的おさらい

つまり、概要配列と同じ塩基の領域なんて興味ないんです。興味ある領域は、この画面上であえて言えば①になります。概要側はGで、マップされたリードは2本(depth = 2)。1つはGで概要側と同じ、そしてもう1つはTで異なるからです



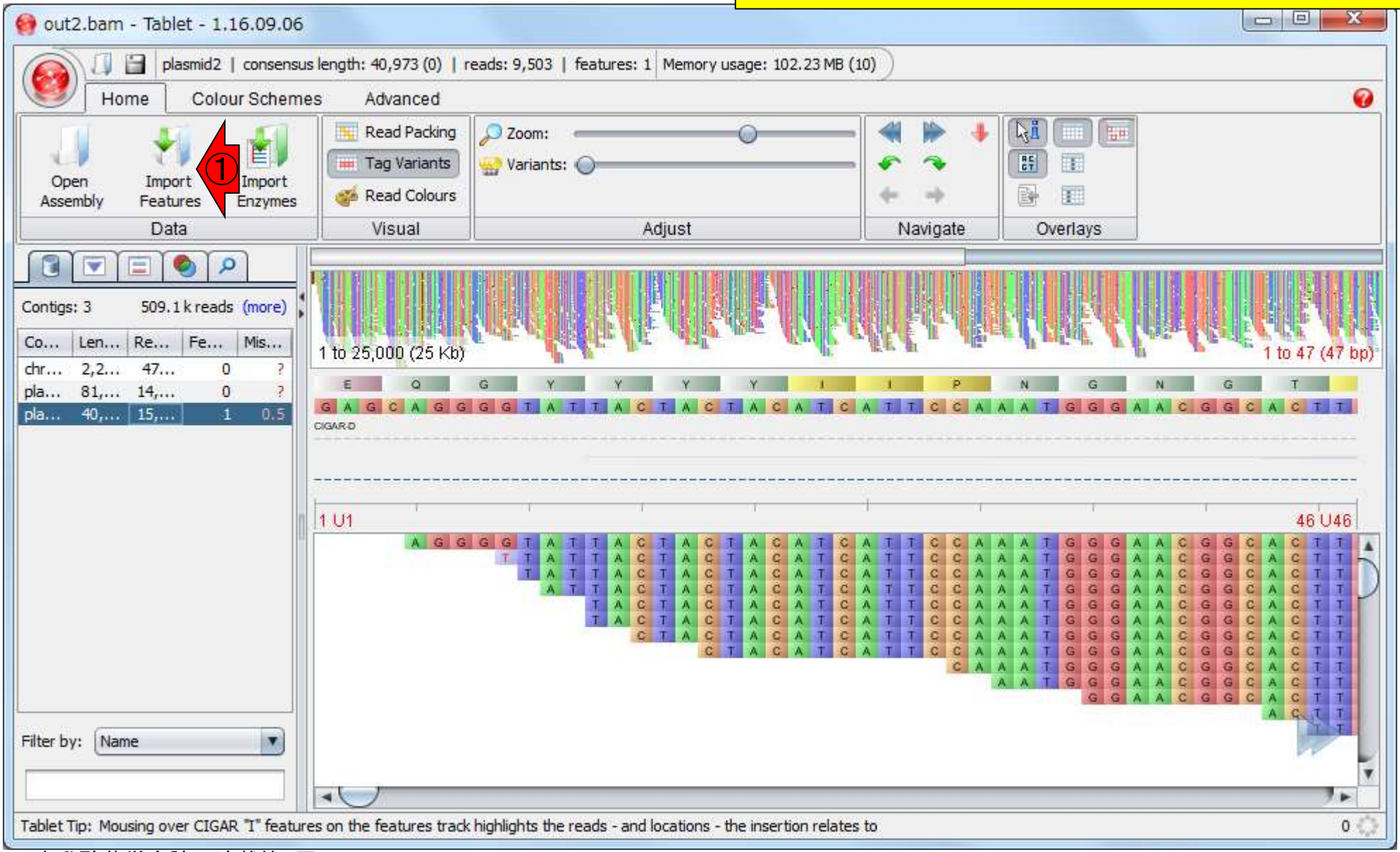
W19-5：目的おさら

しかし、この領域（塩基）の場合はdepth = 2なので、本当はTかもしれないけどGでいいのかもしれないと悩みます。結論としては概要側のGをTに修正しません。理由はdepth = 2で低い（つまり信頼性が低い）からです。例えばこの領域に沢山のリードがマップされていて、そのほとんどがTなら、概要側のGをTに修正します



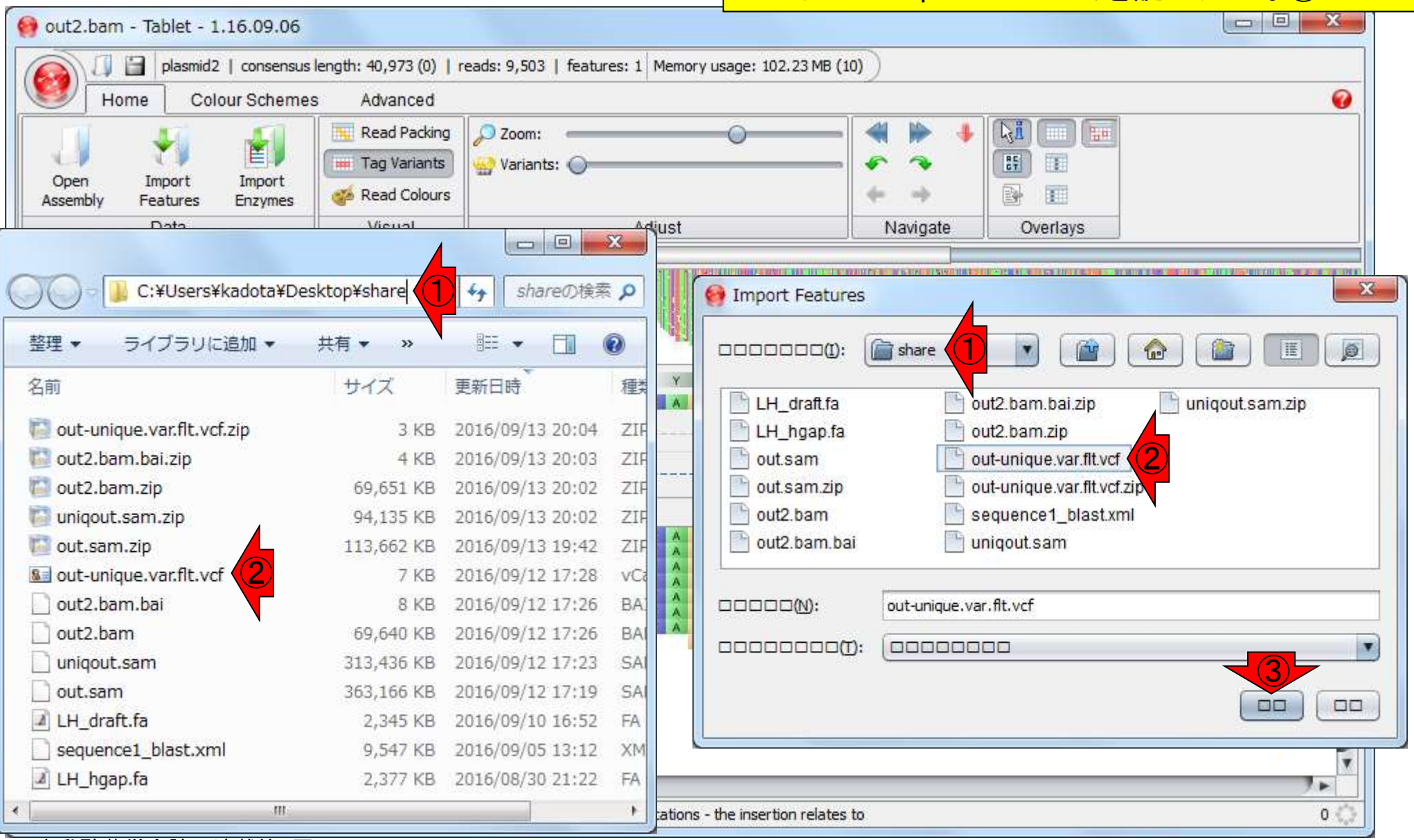
W19-6：VCF

次に読み込むVCFファイル(out-unique.var.flt.vcf)は、一定の信頼性のある概要配列と異なる部分の情報を保持している。①Import Features



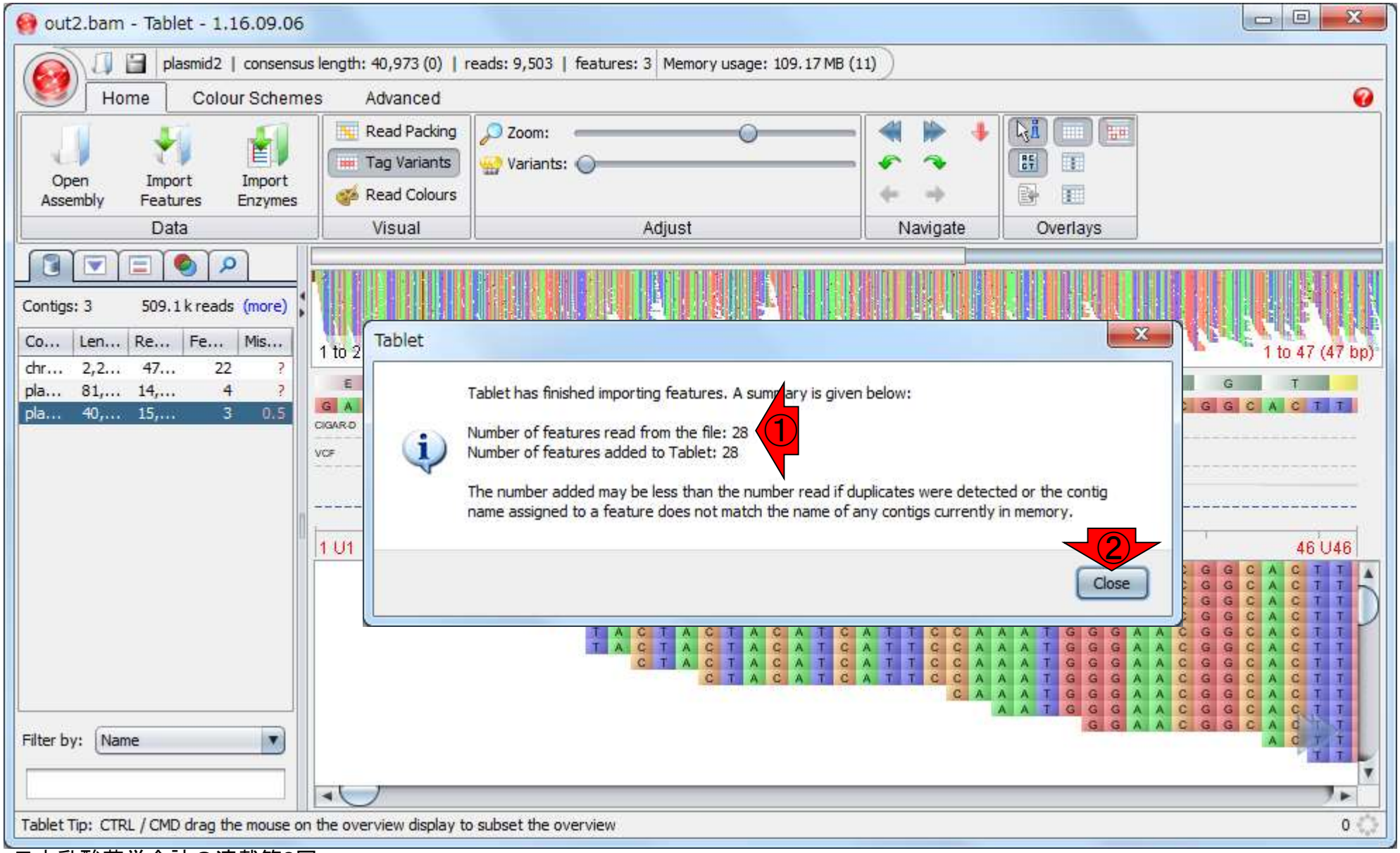
W19-6：VCF

文字化けしているようで見づらいが、どうにかして①（ここでは）共有フォルダ上にある、②VCFファイル(out-unique.var.flt.vcf)を読み込む。③ここ



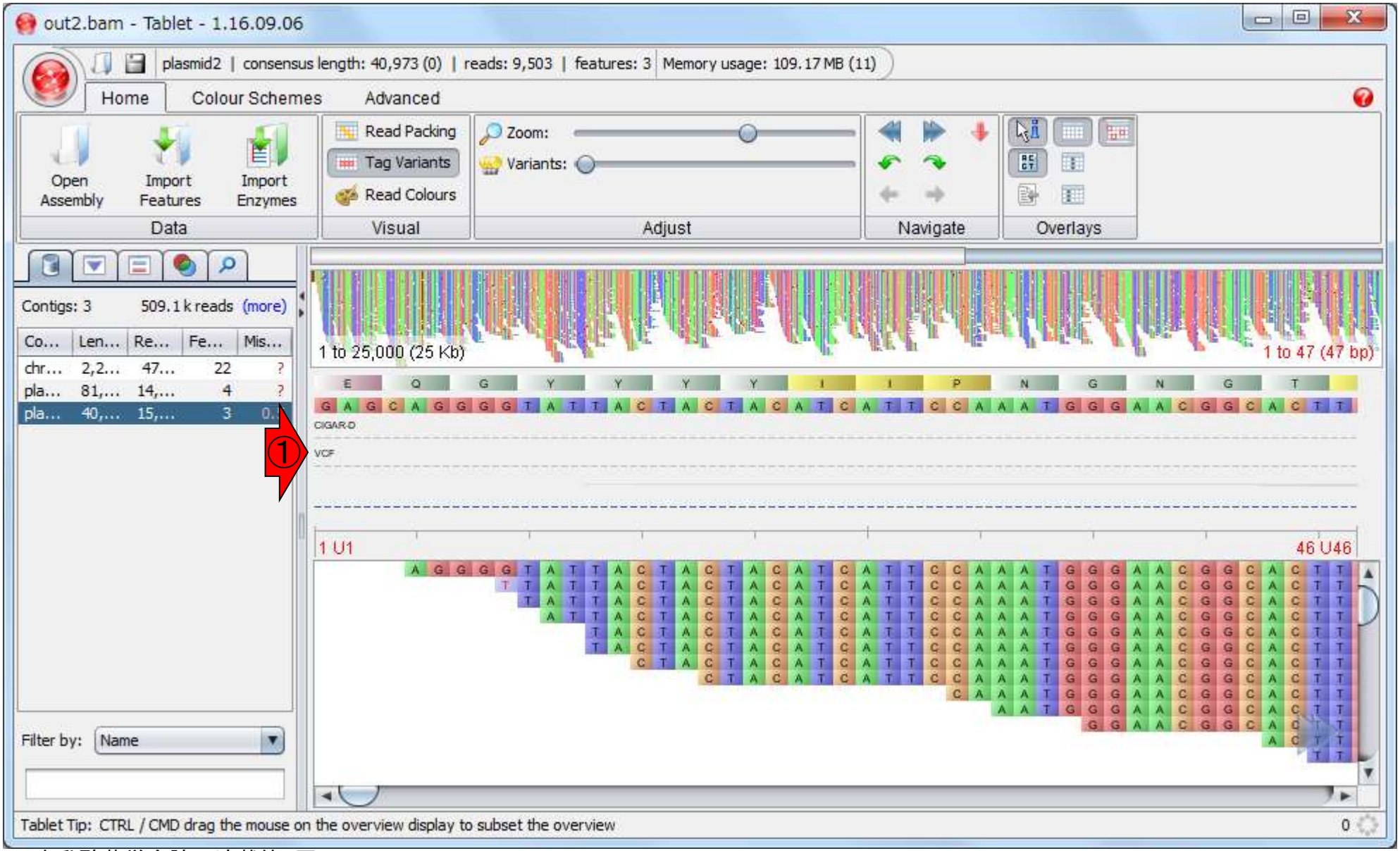
W19-6：VCF

①28個のfeatureがあるようだ。これは変異数に相当(W18-2)。②Close



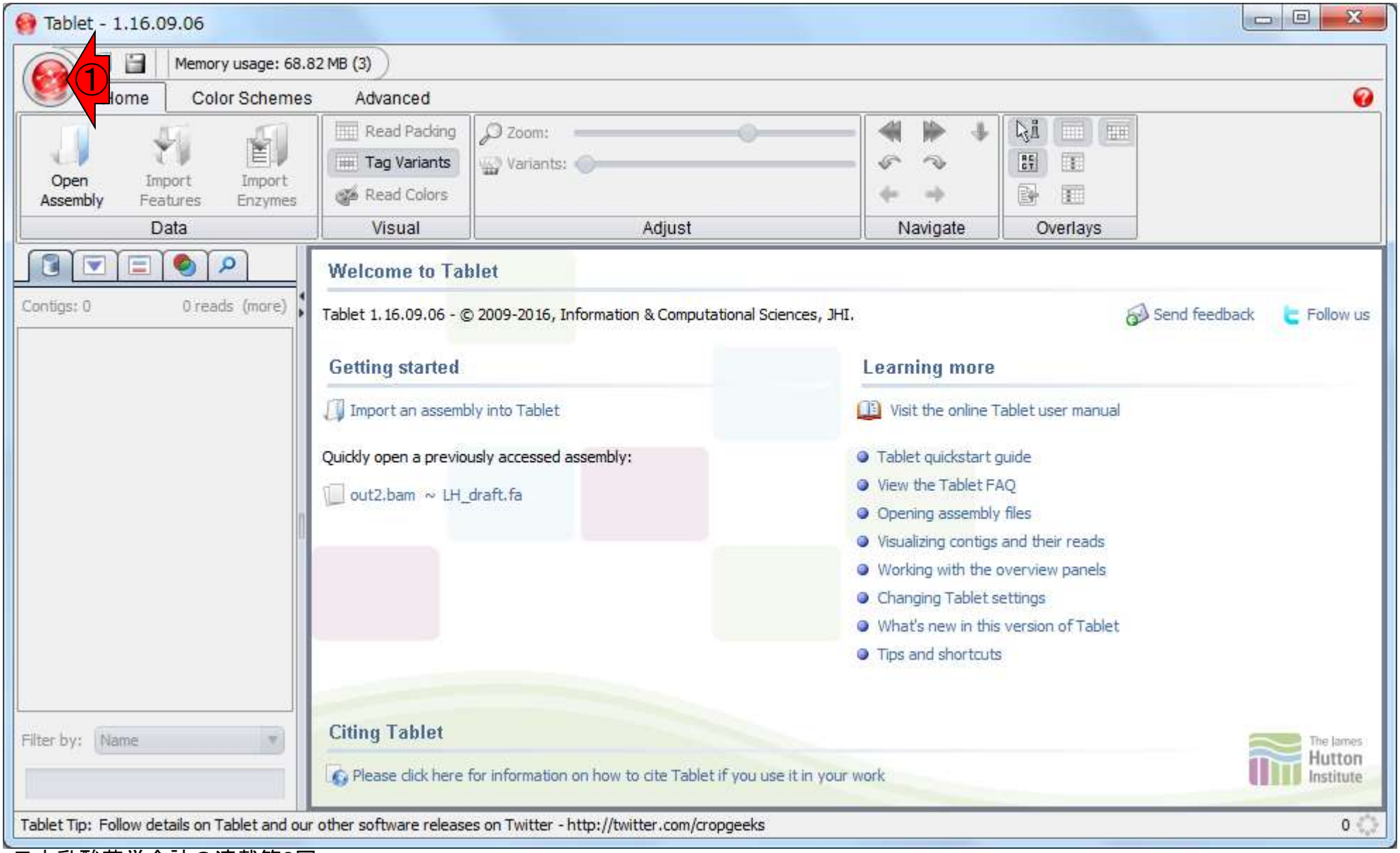
W19-6：VCF

①この部分にVCFの情報が追加されていることがわかります



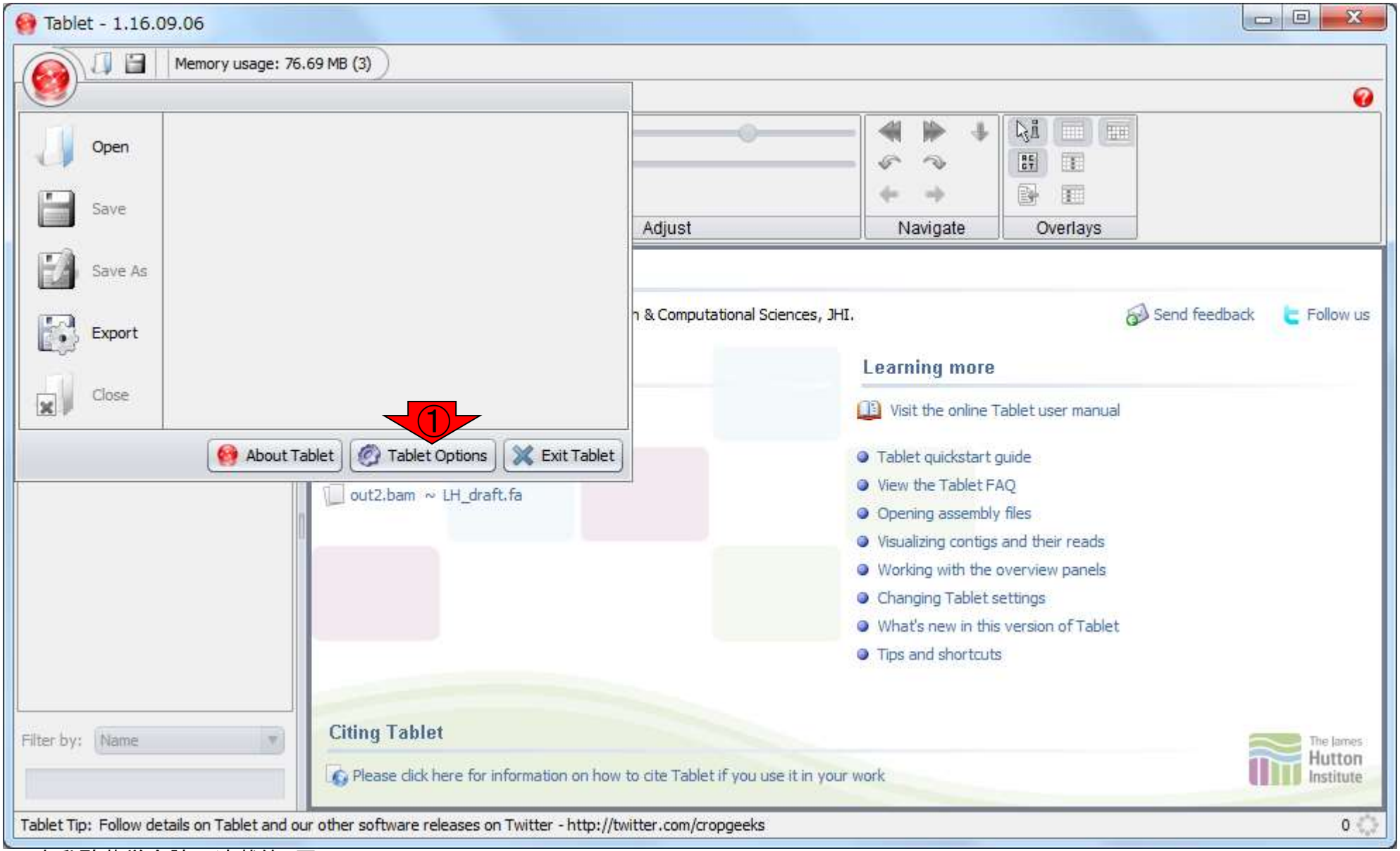
W19-7：文字化け回避策

Tipsとして文字化け回避策を示す。①をクリック



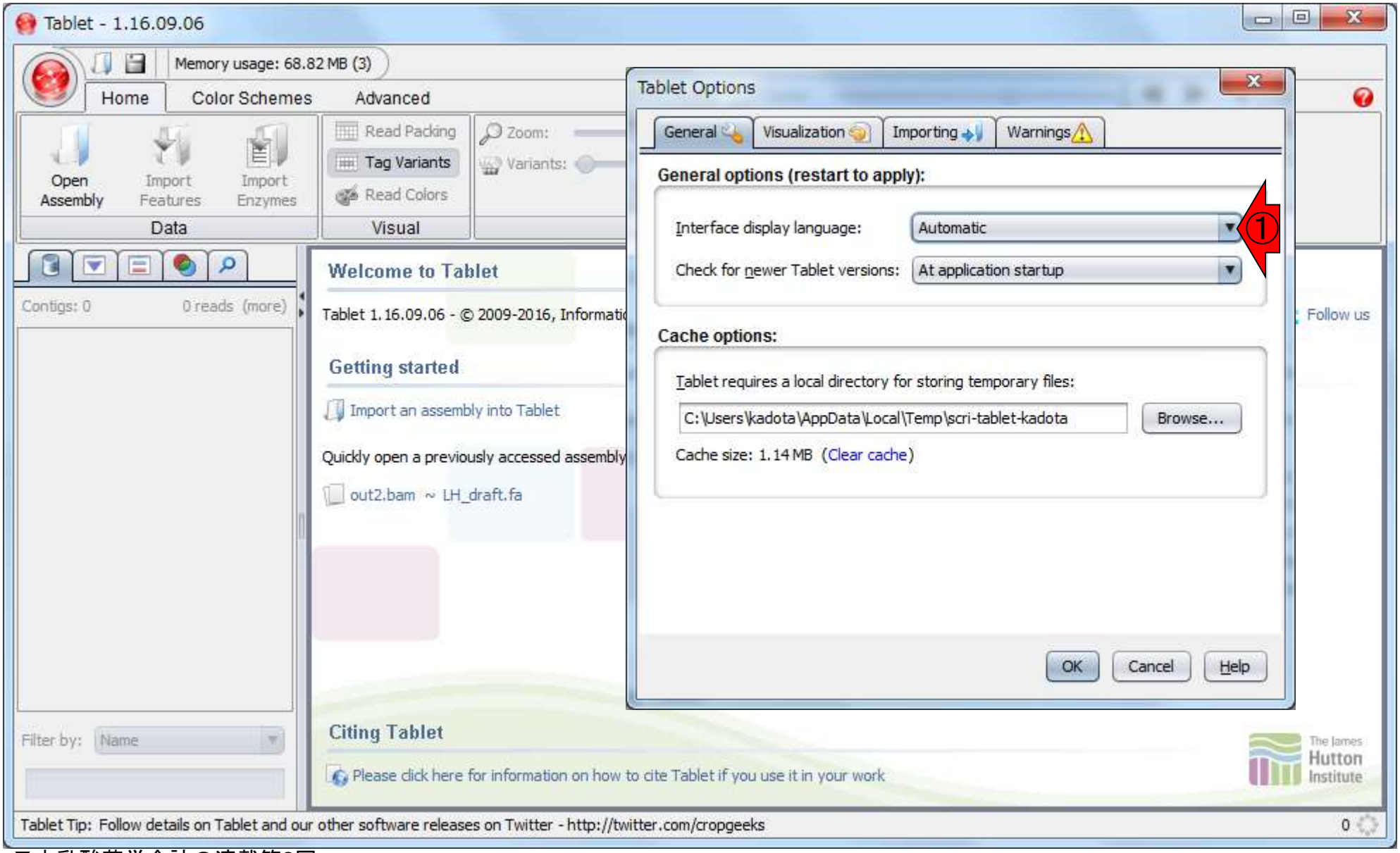
こんな感じになります。①Tablet Options

W19-7：文字化け回避策



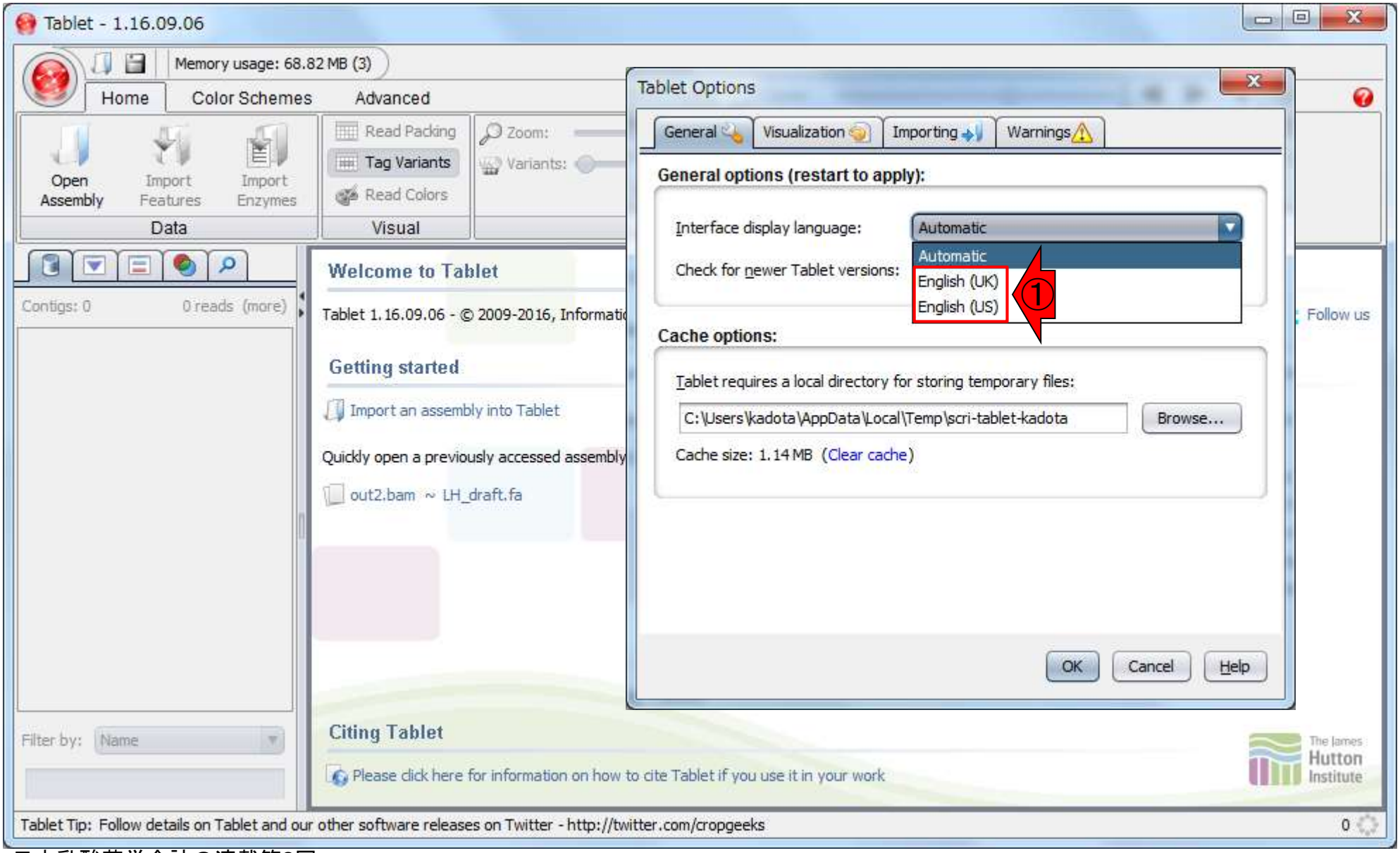
①Automaticを…

W19-7：文字化け回避策



W19-7：文字化け回避策

①English(UK)またはEnglish(US)にすればうまくいくかも…



W20-1：plasmid2

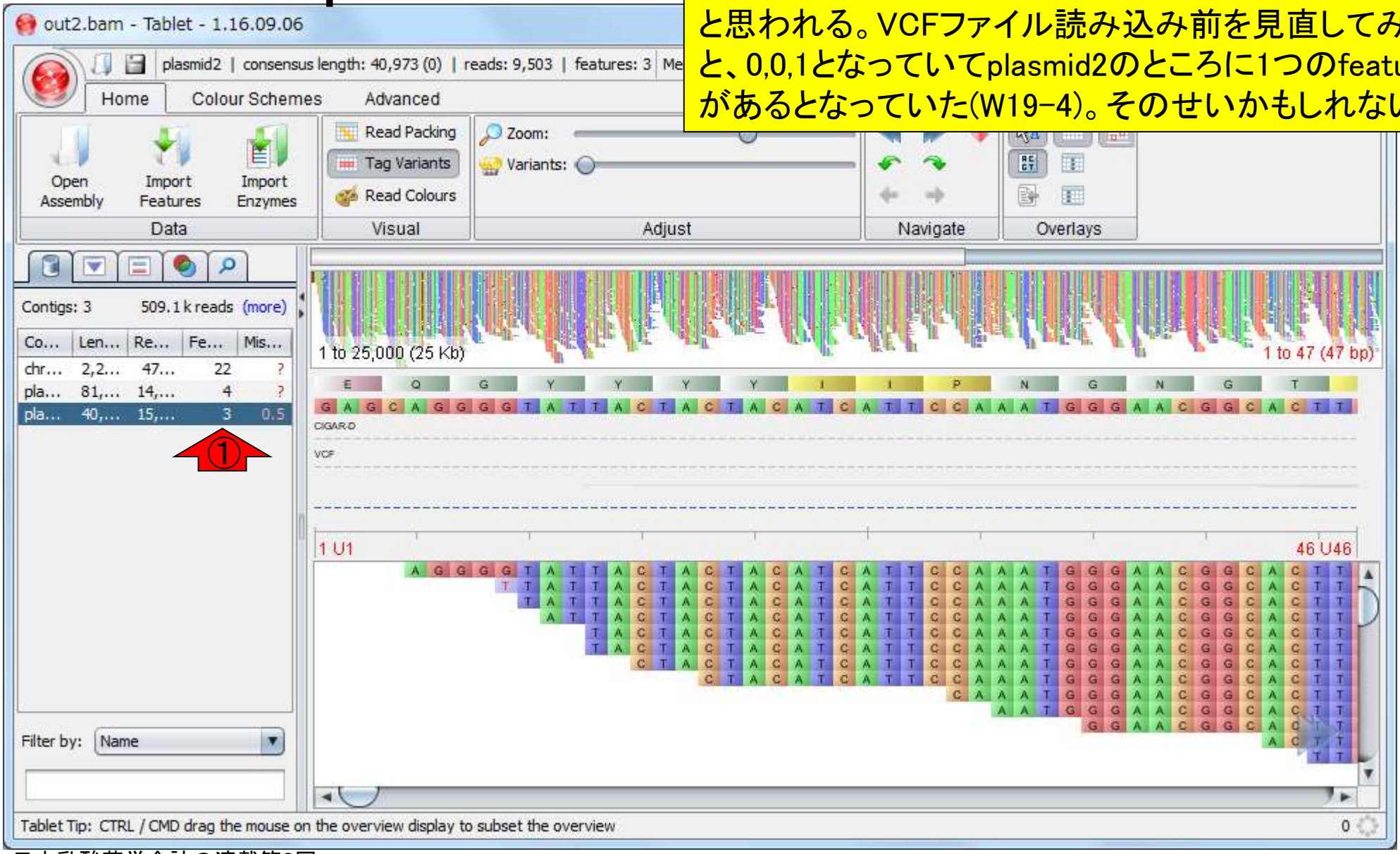
VCFファイルの27行目以降を、Excelで眺めたものを再掲(W18-2)。①plasmid2上には、2か所変異があるようなので2,569番目と37,805番目をTabletで眺める

	A	B	C	D	E	F	G	H	I	J
27	#CHROM	POS	ID	REF	ALT	QUAL	FILTER	INFO	FORMAT	/home/w3pipeline/refdata/
28	chromosome	141598	.	A	G	17.1	.	DP=7;VDB=0.0077;AF1=0.5;AC1	GT:PL:GQ	0/1:47,0,138:50
29	chromosome	242819	.	T	C	20	.	DP=9;VDB=0.0111;AF1=0.5;AC1	GT:PL:GQ	0/1:50,0,139:53
30	chromosome	359141	.	T	C	4.13	.	DP=14;VDB=0.0037;AF1=0.4998	GT:PL:GQ	0/1:32,0,187:31
31	chromosome	463882	.	T	G	6.2	.	DP=13;VDB=0.0074;AF1=0.4999	GT:PL:GQ	0/1:35,0,216:36
32	chromosome	660031	.	C	T	8.64	.	DP=11;VDB=0.0063;AF1=0.5;AC	GT:PL:GQ	0/1:38,0,197:40
33	chromosome	792963	.	T	C	4.13	.	DP=14;VDB=0.0153;AF1=0.4998	GT:PL:GQ	0/1:32,0,200:31
34	chromosome	887422	.	G	A	10.4	.	DP=11;VDB=0.0070;AF1=0.5;AC	GT:PL:GQ	0/1:40,0,188:42
35	chromosome	913470	.	T	C	14.2	.	DP=9;VDB=0.0056;AF1=0.5;AC1	GT:PL:GQ	0/1:44,0,144:47
36	chromosome	965171	.	T	C	17.1	.	DP=20;VDB=0.0071;AF1=0.5;AC	GT:PL:GQ	0/1:47,0,249:50
37	chromosome	1134950	.	A	G	7.8	.	DP=23;VDB=0.0076;AF1=0.5;AC	GT:PL:GQ	0/1:37,0,255:39
38	chromosome	1160048	.	A	G	29	.	DP=17;VDB=0.0054;AF1=0.5;AC	GT:PL:GQ	0/1:59,0,255:62
39	chromosome	1160095	.	A	G	19.1	.	DP=10;VDB=0.0033;AF1=0.5;AC	GT:PL:GQ	0/1:49,0,191:52
40	chromosome	1240248	.	GCCCC	GCCCCC	214	.	INDEL;DP=48;VDB=0.0137;AF1=	GT:PL:GQ	1/1:255,128,0:99
41	chromosome	1455552	.	CAAAA	CAAAAA	214	.	INDEL;DP=58;VDB=0.0164;AF1=	GT:PL:GQ	1/1:255,130,0:99
42	chromosome	1475721	.	T	C	52	.	DP=19;VDB=0.0021;AF1=0.5;AC	GT:PL:GQ	0/1:82,0,246:85
43	chromosome	1752680	.	G	A	29	.	DP=24;VDB=0.0129;AF1=0.5;AC	GT:PL:GQ	0/1:59,0,255:62
44	chromosome	1941380	.	T	C	7.8	.	DP=13;VDB=0.0046;AF1=0.5;AC	GT:PL:GQ	0/1:37,0,214:39
45	chromosome	2052955	.	A	G	12.3	.	DP=8;VDB=0.0084;AF1=0.5;AC1	GT:PL:GQ	0/1:42,0,103:44
46	chromosome	2054084	.	TC	T	13.7	.	INDEL;DP=8;VDB=0.0009;AF1=0	GT:PL:GQ	0/1:51,0,174:54
47	chromosome	2056279	.	A	G	24	.	DP=6;VDB=0.0057;AF1=0.5;AC1	GT:PL:GQ	0/1:54,0,113:57
48	chromosome	2057564	.	T	C	8.64	.	DP=10;VDB=0.0041;AF1=0.5;AC	GT:PL:GQ	0/1:38,0,129:40
49	chromosome	2058145	.	C	T,G	108	.	DP=14;VDB=0.0147;AF1=1;AC1=	GT:PL:GQ	1/1:141,39,0,141,28,138:75
50	plasmid1	37177	.	G	T	54	.	DP=49;VDB=0.0087;AF1=0.5;AC	GT:PL:GQ	0/1:84,0,255:87
51	plasmid1	66609	.	A	G	225	.	DP=55;VDB=0.0060;AF1=0.5;AC	GT:PL:GQ	0/1:255,0,255:99
52	plasmid1	66637	.	C	T	225	.	DP=71;VDB=0.0141;AF1=0.5;AC	GT:PL:GQ	0/1:255,0,249:99
53	plasmid1	66668	.	C	T	225	.	DP=80;VDB=0.0153;AF1=0.5;AC	GT:PL:GQ	0/1:255,0,255:99
54	plasmid2	2569	.	CAAAAA	CAAAA	177	.	INDEL;DP=110;VDB=0.0166;AF1	GT:PL:GQ	1/1:218,255,0:99
55	plasmid2	37805	.	TAAAAA	TAAAA	104	.	INDEL;DP=31;VDB=0.0058;AF1=	GT:PL:GQ	1/1:145,93,0:99

①

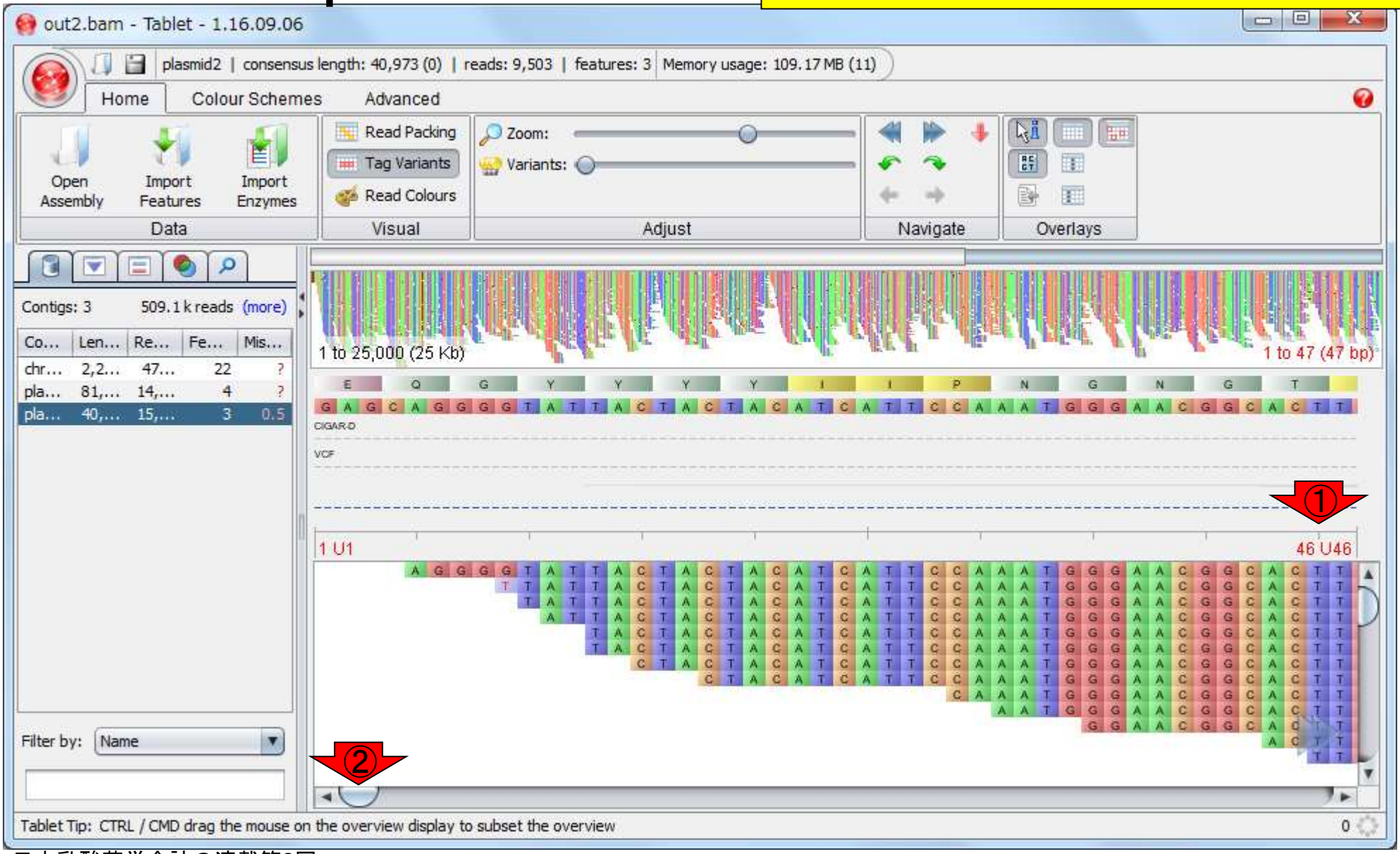
W20-1：plasmid2

①VCFファイル読み込み時に28 featuresと書いてあったが、ここでは(22+4+3) =29個となっている。plasmid2が2個ではなく3個となっているところが変だと思われる。VCFファイル読み込み前を見直してみると、0,0,1となっていてplasmid2のところに1つのfeatureがあるとなっていた(W19-4)。そのせいかもしれない



W20-1 : plasmid2

まずは1つめの変異である2,569番目の塩基を見に行く。①ここが概要側の塩基番号に相当するので、ここを頼りに②を右スクロールしていく



①1つめの変異である2,569番目を発見

W20-1：plasmid2

W20-1：plasmid2

①リファレンス、CAAAAA。②ALT列の変異側（つまりマップする側のIlluminaリード配列）は、CAAAA。確かにマッピング結果はそうなっている。③DP=110なので、2,569番目のポジションには、110個のリードがマップされた結果だということがわかる

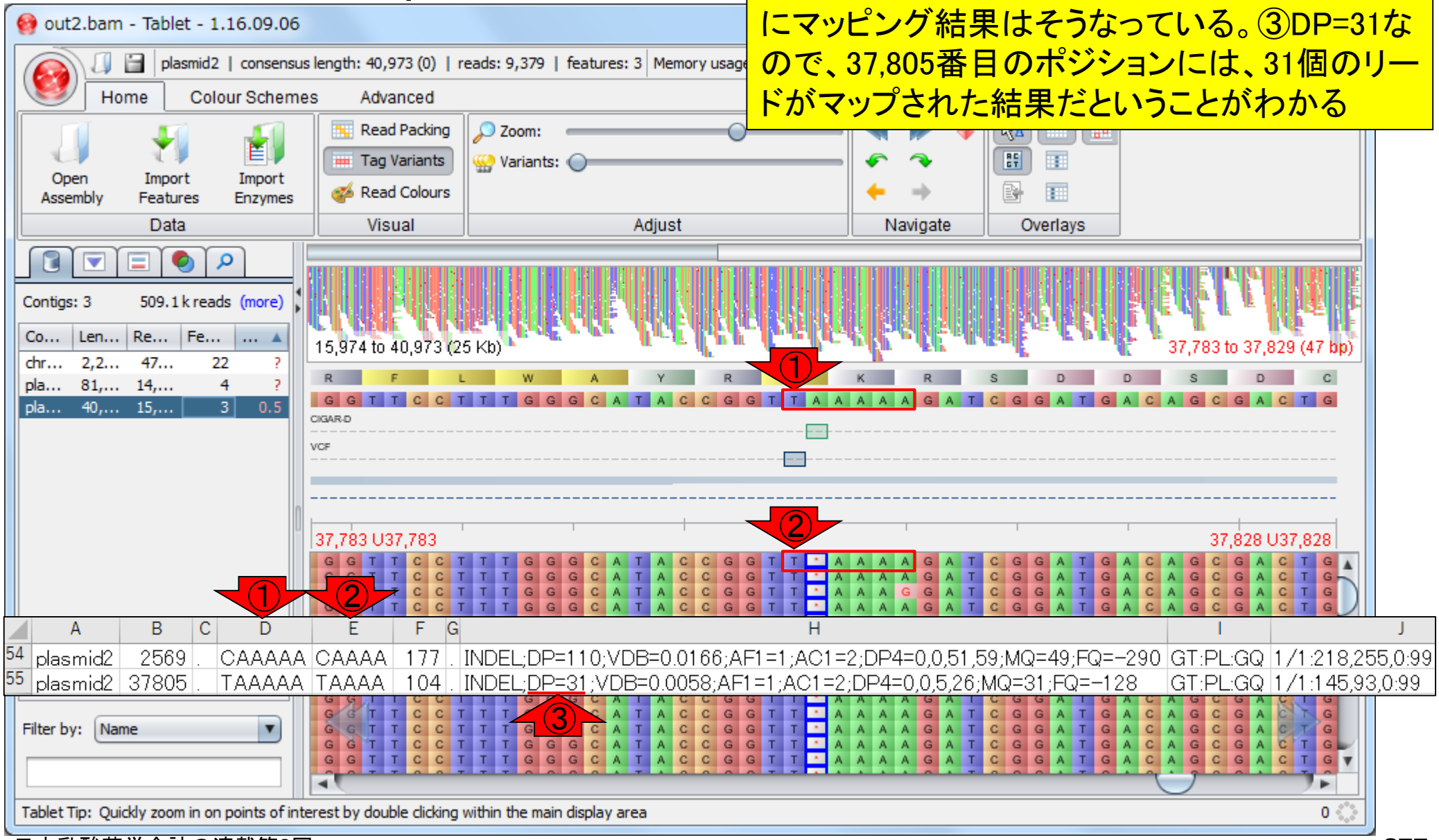
	A	B	C	D	E	F	G	H	I	J
54	plasmid2	2569	.	CAAAAA	CAAAA	177	.	INDEL;DP=110;VDB=0.0166;AF1=1;AC1=2;DP4=0,0,51,59;MQ=49;FQ=-290	GT:PL:GQ	1/1:218,255,0:99
55	plasmid2	37805	.	TAAAAA	TAAAA	104	.	INDEL;DP=[illegible];VDB=0.0058;AF1=1;AC1=2;DP4=0,0,5,26;MQ=31;FQ=-128	GT:PL:GQ	1/1:145,93,0:99

W20-1 : plasmid2

①下のほうに移動させて全体を眺めたところ。確かに概要配列のCAAAAAから、Aを1つ除去してCAAAAとしたほうがよさそうと判断できる

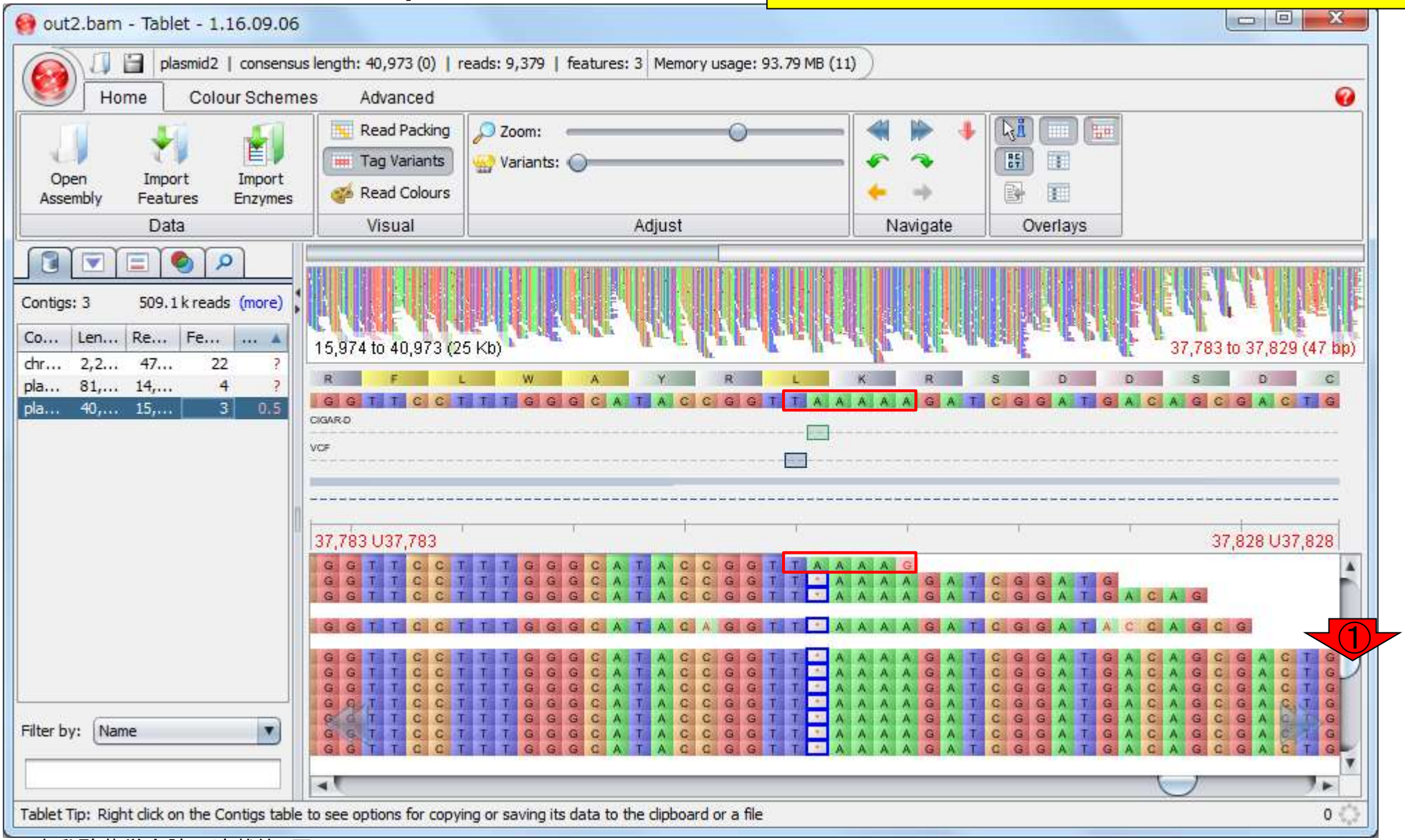
W20-2：37,805番目

plasmid2中の2つめの変異を確認。①リファレンス、TAAAAA。②ALT列の変異側（つまりマップする側のIlluminaリード配列）は、TAAAA。確かにマッピング結果はそうなっている。③DP=31なので、37,805番目のポジションには、31個のリードがマップされた結果だということがわかる



W20-2：37,805番目

①下のほうに移動させて全体を眺めたところ。確かに概要配列のTAAAAAから、Aを1つ除去してTAAAAとしたほうがよさそうと判断できる



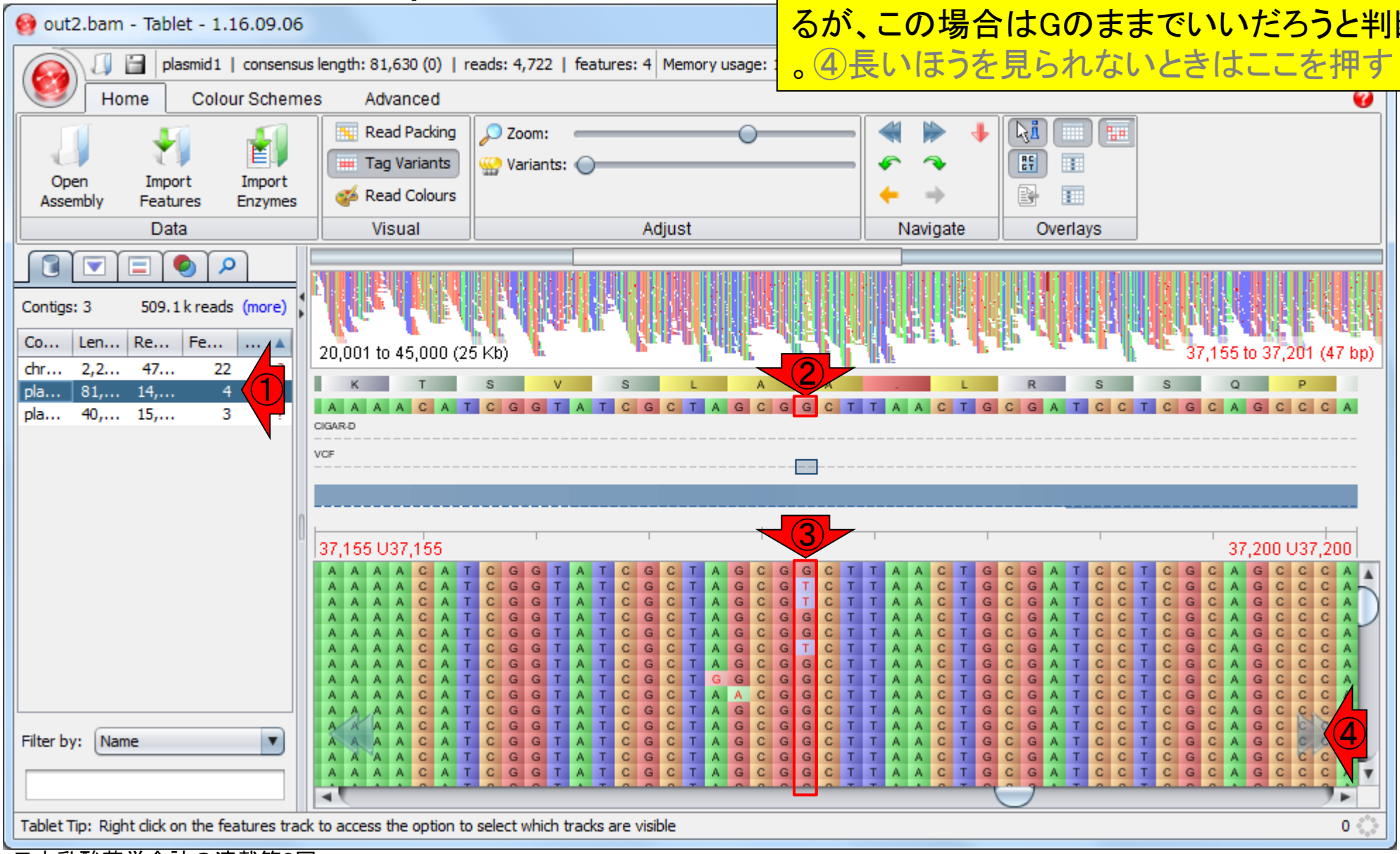
次は、①plasmid1上の、4か所の変異をチェック。まずは②37,177番目

W21-1：plasmid1

	A	B	C	D	E	F	G	H	I	J
27	#CHROM	POS	ID	REF	ALT	QUAL	FILTER	INFO	FORMAT	/home/w3pipeline/refdata/
28	chromosome	141598	.	A	G	17.1	.	DP=7;VDB=0.0077;AF1=0.5;AC1	GT:PL:GQ	0/1:47,0,138:50
29	chromosome	242819	.	T	C	20	.	DP=9;VDB=0.0111;AF1=0.5;AC1	GT:PL:GQ	0/1:50,0,139:53
30	chromosome	359141	.	T	C	4.13	.	DP=14;VDB=0.0037;AF1=0.4998	GT:PL:GQ	0/1:32,0,187:31
31	chromosome	463882	.	T	G	6.2	.	DP=13;VDB=0.0074;AF1=0.4999	GT:PL:GQ	0/1:35,0,216:36
32	chromosome	660031	.	C	T	8.64	.	DP=11;VDB=0.0063;AF1=0.5;AC	GT:PL:GQ	0/1:38,0,197:40
33	chromosome	792963	.	T	C	4.13	.	DP=14;VDB=0.0153;AF1=0.4998	GT:PL:GQ	0/1:32,0,200:31
34	chromosome	887422	.	G	A	10.4	.	DP=11;VDB=0.0070;AF1=0.5;AC	GT:PL:GQ	0/1:40,0,188:42
35	chromosome	913470	.	T	C	14.2	.	DP=9;VDB=0.0056;AF1=0.5;AC1	GT:PL:GQ	0/1:44,0,144:47
36	chromosome	965171	.	T	C	17.1	.	DP=20;VDB=0.0071;AF1=0.5;AC	GT:PL:GQ	0/1:47,0,249:50
37	chromosome	1134950	.	A	G	7.8	.	DP=23;VDB=0.0076;AF1=0.5;AC	GT:PL:GQ	0/1:37,0,255:39
38	chromosome	1160048	.	A	G	29	.	DP=17;VDB=0.0054;AF1=0.5;AC	GT:PL:GQ	0/1:59,0,255:62
39	chromosome	1160095	.	A	G	19.1	.	DP=10;VDB=0.0033;AF1=0.5;AC	GT:PL:GQ	0/1:49,0,191:52
40	chromosome	1240248	.	GCCCC	GCCCCC	214	.	INDEL;DP=48;VDB=0.0137;AF1=	GT:PL:GQ	1/1:255,128,0:99
41	chromosome	1455552	.	CAAAA	CAAAAA	214	.	INDEL;DP=58;VDB=0.0164;AF1=	GT:PL:GQ	1/1:255,130,0:99
42	chromosome	1475721	.	T	C	52	.	DP=19;VDB=0.0021;AF1=0.5;AC	GT:PL:GQ	0/1:82,0,246:85
43	chromosome	1752680	.	G	A	29	.	DP=24;VDB=0.0129;AF1=0.5;AC	GT:PL:GQ	0/1:59,0,255:62
44	chromosome	1941380	.	T	C	7.8	.	DP=13;VDB=0.0046;AF1=0.5;AC	GT:PL:GQ	0/1:37,0,214:39
45	chromosome	2052955	.	A	G	12.3	.	DP=8;VDB=0.0084;AF1=0.5;AC1	GT:PL:GQ	0/1:42,0,103:44
46	chromosome	2054084	.	TC	T	13.7	.	INDEL;DP=8;VDB=0.0009;AF1=0	GT:PL:GQ	0/1:51,0,174:54
47	chromosome	2056279	.	A	G	24	.	DP=6;VDB=0.0057;AF1=0.5;AC1	GT:PL:GQ	0/1:54,0,113:57
48	chromosome	2057564	.	T	C	8.64	.	DP=10;VDB=0.0041;AF1=0.5;AC	GT:PL:GQ	0/1:38,0,129:40
49	chromosome	2058145	.	C	T,G	108	.	DP=14;VDB=0.0147;AF1=1;AC1=	GT:PL:GQ	1/1:141,39,0,141,28,138:75
50	plasmid1	37177	.	G	T	54	.	DP=49;VDB=0.0087;AF1=0.5;AC	GT:PL:GQ	0/1:84,0,255:87
51	plasmid1	66609	.	A	G	225	.	DP=55;VDB=0.0060;AF1=0.5;AC	GT:PL:GQ	0/1:255,0,255:99
52	plasmid1	66637	.	C	T	225	.	DP=71;VDB=0.0141;AF1=0.5;AC	GT:PL:GQ	0/1:255,0,249:99
53	plasmid1	66668	.	C	T	225	.	DP=80;VDB=0.0153;AF1=0.5;AC	GT:PL:GQ	0/1:255,0,255:99
54	plasmid2	2569	.	CAAAAA	CAAAA	177	.	INDEL;DP=110;VDB=0.0166;AF1	GT:PL:GQ	1/1:218,255,0:99
55	plasmid2	37805	.	TAAAAA	TAAAA	104	.	INDEL;DP=31;VDB=0.0058;AF1=	GT:PL:GQ	1/1:145,93,0:99

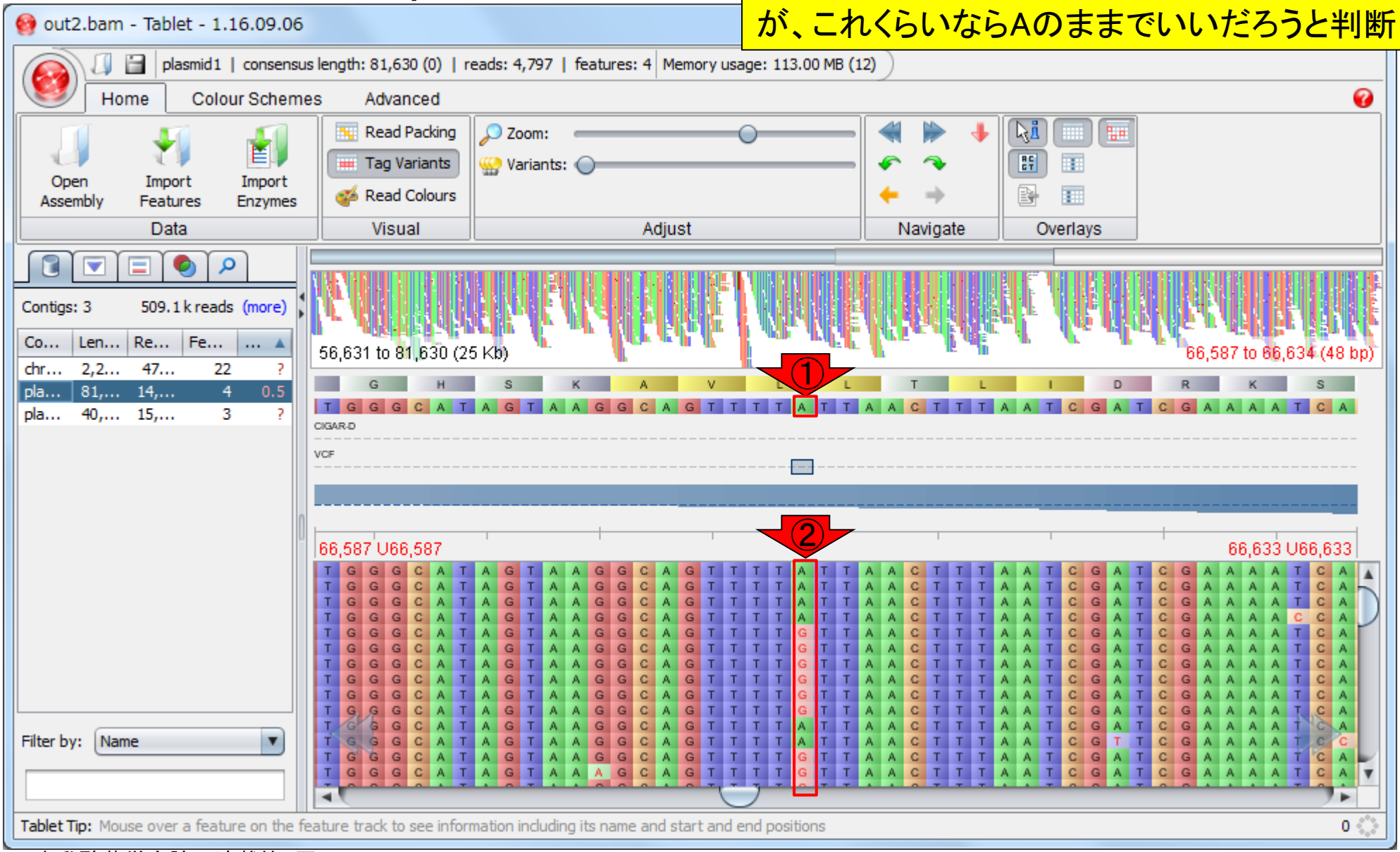
W21-1:37,177番目

①plasmid1に切り替え。②リファレンスはG。③ALT列の変異側（つまりマップする側のIlluminaリード配列）はT。そういうのも確かにあるが、この場合はGのままでいいだろうと判断。④長いほうを見られないときはここを押す



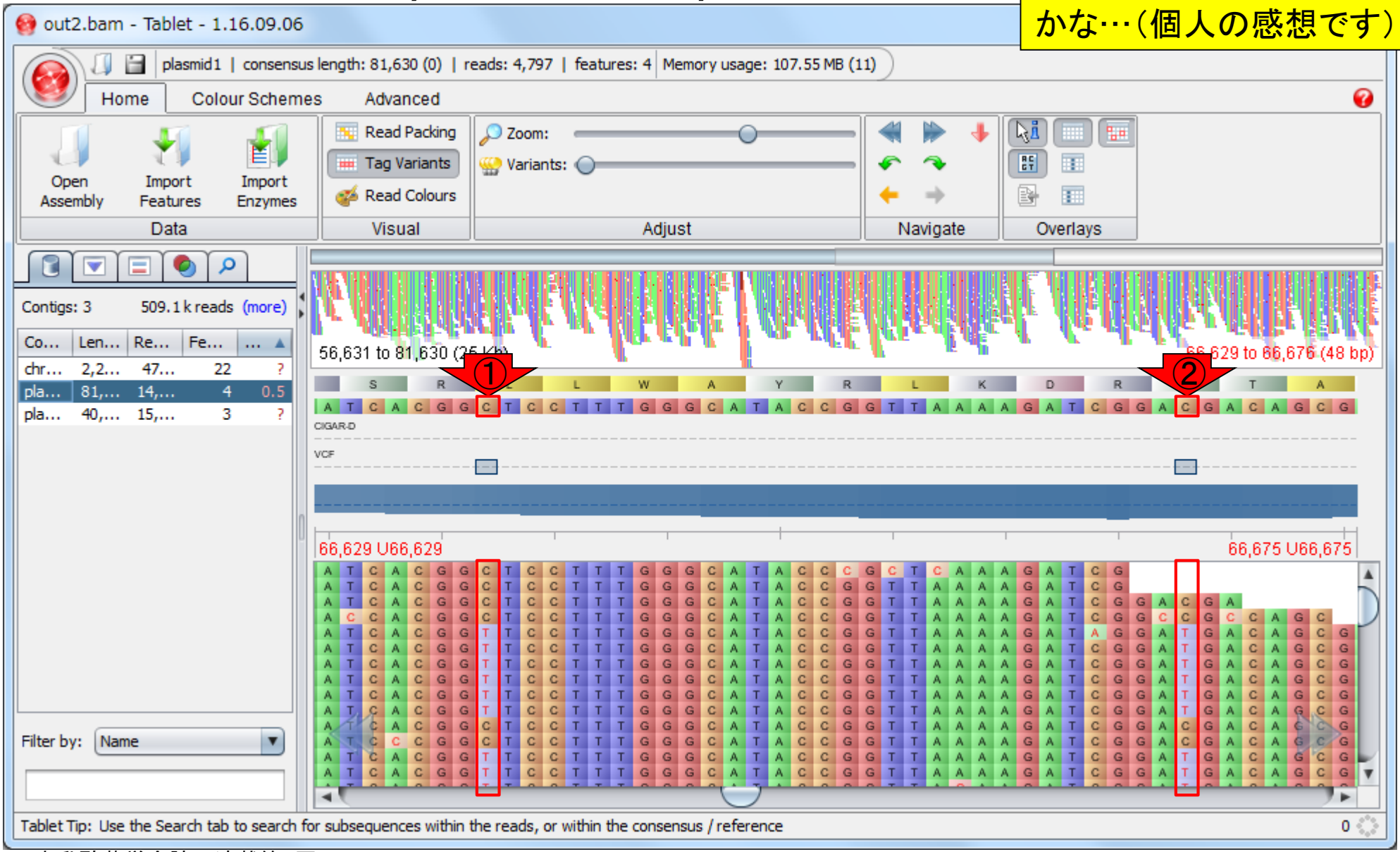
W21-2：66,609番目

66,609番目の塩基を表示。①リファレンスはA。②ALT列の変異側（つまりマップする側のIllumina リード配列）はG。Gが全体的にAに比べて多いが、これくらいならAのままでいいだろうと判断



W21-3：66,637と66,668番目

①66,637番目と②66,668番目の塩基を表示。このあたりも概要側を変えなくてもいいかな…（個人の感想です）



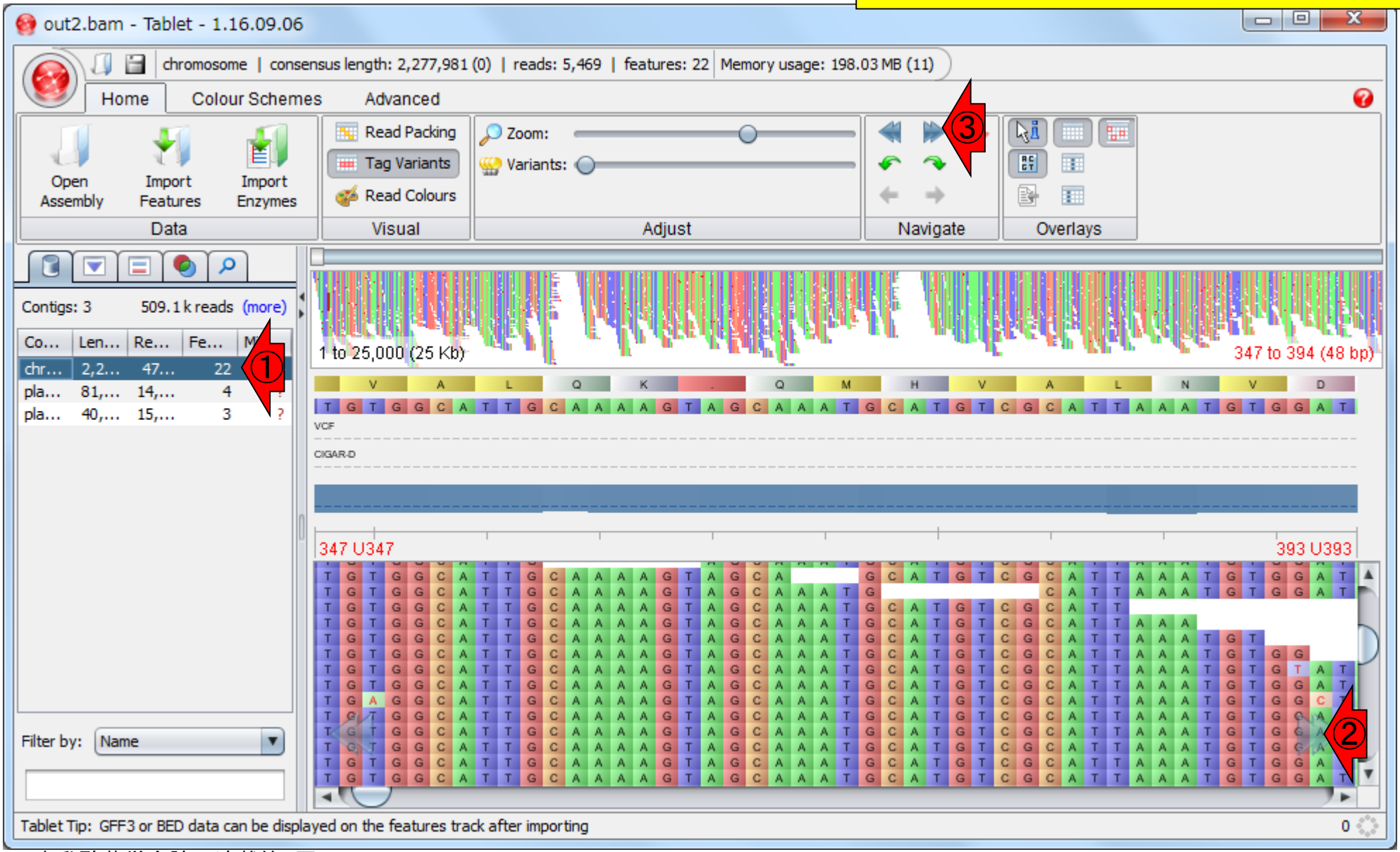
W22-1：chromosome

今度はchromosomeを確認。
とりあえず①2,054,084番目

	A	B	C	D	E	F	G	H	I	J
27	#CHROM	POS	ID	REF	ALT	QUAL	FILTER	INFO	FORMAT	/home/w3pipeline/refdata/
28	chromosome	141598	.	A	G	17.1	.	DP=7;VDB=0.0077;AF1=0.5;AC1	GT:PL:GQ	0/1:47,0,138:50
29	chromosome	242819	.	T	C	20	.	DP=9;VDB=0.0111;AF1=0.5;AC1	GT:PL:GQ	0/1:50,0,139:53
30	chromosome	359141	.	T	C	4.13	.	DP=14;VDB=0.0037;AF1=0.4998	GT:PL:GQ	0/1:32,0,187:31
31	chromosome	463882	.	T	G	6.2	.	DP=13;VDB=0.0074;AF1=0.4999	GT:PL:GQ	0/1:35,0,216:36
32	chromosome	660031	.	C	T	8.64	.	DP=11;VDB=0.0063;AF1=0.5;AC	GT:PL:GQ	0/1:38,0,197:40
33	chromosome	792963	.	T	C	4.13	.	DP=14;VDB=0.0153;AF1=0.4998	GT:PL:GQ	0/1:32,0,200:31
34	chromosome	887422	.	G	A	10.4	.	DP=11;VDB=0.0070;AF1=0.5;AC	GT:PL:GQ	0/1:40,0,188:42
35	chromosome	913470	.	T	C	14.2	.	DP=9;VDB=0.0056;AF1=0.5;AC1	GT:PL:GQ	0/1:44,0,144:47
36	chromosome	965171	.	T	C	17.1	.	DP=20;VDB=0.0071;AF1=0.5;AC	GT:PL:GQ	0/1:47,0,249:50
37	chromosome	1134950	.	A	G	7.8	.	DP=23;VDB=0.0076;AF1=0.5;AC	GT:PL:GQ	0/1:37,0,255:39
38	chromosome	1160048	.	A	G	29	.	DP=17;VDB=0.0054;AF1=0.5;AC	GT:PL:GQ	0/1:59,0,255:62
39	chromosome	1160095	.	A	G	19.1	.	DP=10;VDB=0.0033;AF1=0.5;AC	GT:PL:GQ	0/1:49,0,191:52
40	chromosome	1240248	.	GCCCC	GCCCCC	214	.	INDEL;DP=48;VDB=0.0137;AF1=	GT:PL:GQ	1/1:255,128,0:99
41	chromosome	1455552	.	CAAAA	CAAAAA	214	.	INDEL;DP=58;VDB=0.0164;AF1=	GT:PL:GQ	1/1:255,130,0:99
42	chromosome	1475721	.	T	C	52	.	DP=19;VDB=0.0021;AF1=0.5;AC	GT:PL:GQ	0/1:82,0,246:85
43	chromosome	1752680	.	G	A	29	.	DP=24;VDB=0.0129;AF1=0.5;AC	GT:PL:GQ	0/1:59,0,255:62
44	chromosome	1941380	.	T	C	7.8	.	DP=13;VDB=0.0046;AF1=0.5;AC	GT:PL:GQ	0/1:37,0,214:39
45	chromosom	2052955	.	A	G	12.3	.	DP=8;VDB=0.0084;AF1=0.5;AC1	GT:PL:GQ	0/1:42,0,103:44
46	chromoso	2054084	.	TC	T	13.7	.	INDEL;DP=8;VDB=0.0009;AF1=0	GT:PL:GQ	0/1:51,0,174:54
47	chromosom	2056279	.	A	G	24	.	DP=6;VDB=0.0057;AF1=0.5;AC1	GT:PL:GQ	0/1:54,0,113:57
48	chromosome	2057564	.	T	C	8.64	.	DP=10;VDB=0.0041;AF1=0.5;AC	GT:PL:GQ	0/1:38,0,129:40
49	chromosome	2058145	.	C	T,G	108	.	DP=14;VDB=0.0147;AF1=1;AC1=	GT:PL:GQ	1/1:141,39,0,141,28,138:75
50	plasmid1	37177	.	G	T	54	.	DP=49;VDB=0.0087;AF1=0.5;AC	GT:PL:GQ	0/1:84,0,255:87
51	plasmid1	66609	.	A	G	225	.	DP=55;VDB=0.0060;AF1=0.5;AC	GT:PL:GQ	0/1:255,0,255:99
52	plasmid1	66637	.	C	T	225	.	DP=71;VDB=0.0141;AF1=0.5;AC	GT:PL:GQ	0/1:255,0,249:99
53	plasmid1	66668	.	C	T	225	.	DP=80;VDB=0.0153;AF1=0.5;AC	GT:PL:GQ	0/1:255,0,255:99
54	plasmid2	2569	.	CAAAAA	CAAAA	177	.	INDEL;DP=110;VDB=0.0166;AF1	GT:PL:GQ	1/1:218,255,0:99
55	plasmid2	37805	.	TAAAAA	TAAAA	104	.	INDEL;DP=31;VDB=0.0058;AF1=	GT:PL:GQ	1/1:145,93,0:99

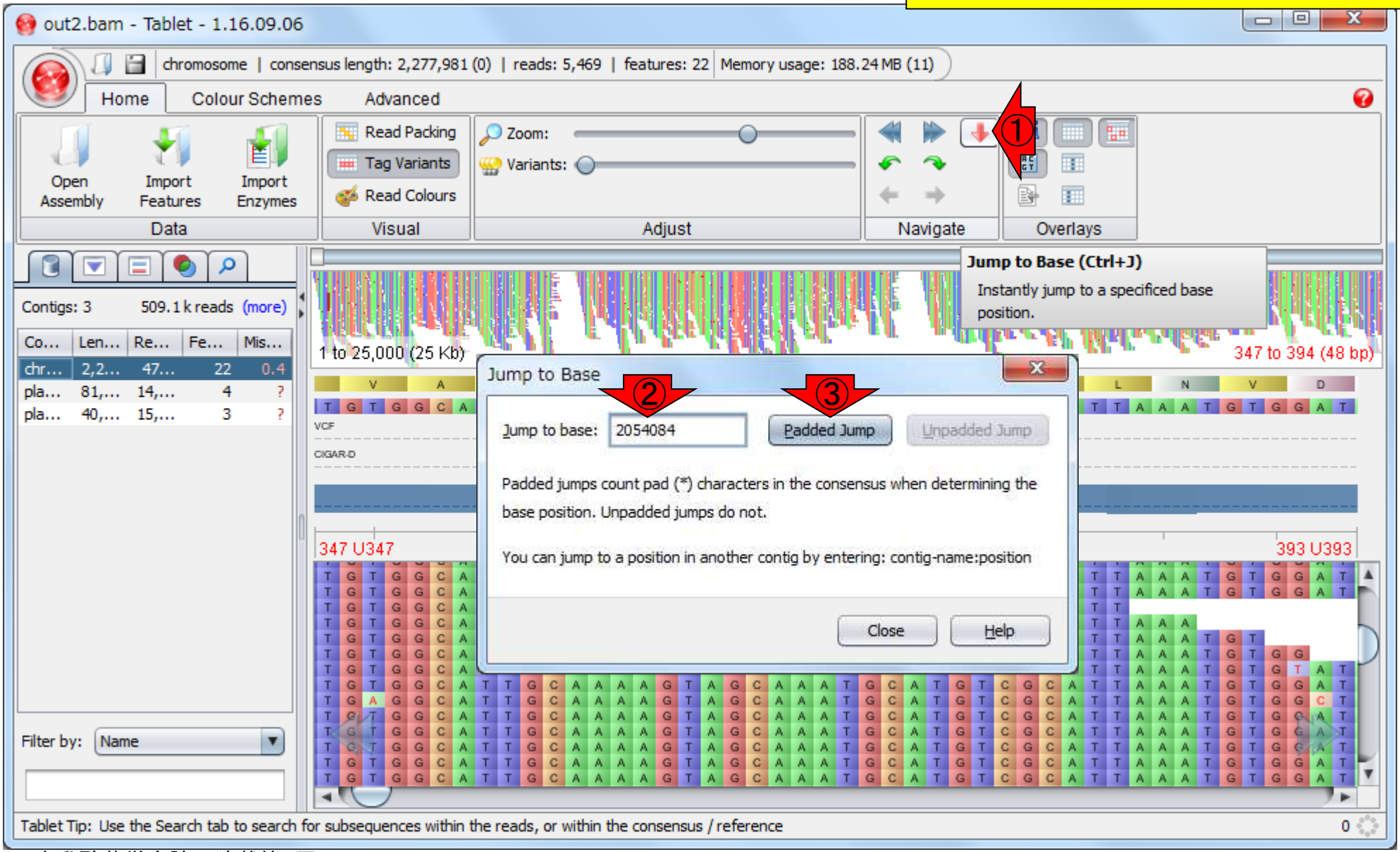
W22-2：2,054,084番目

①chromosomeに変更して、2,054,085番目に移動したい。長すぎるので②または③を押して移動するのは途方に暮れる



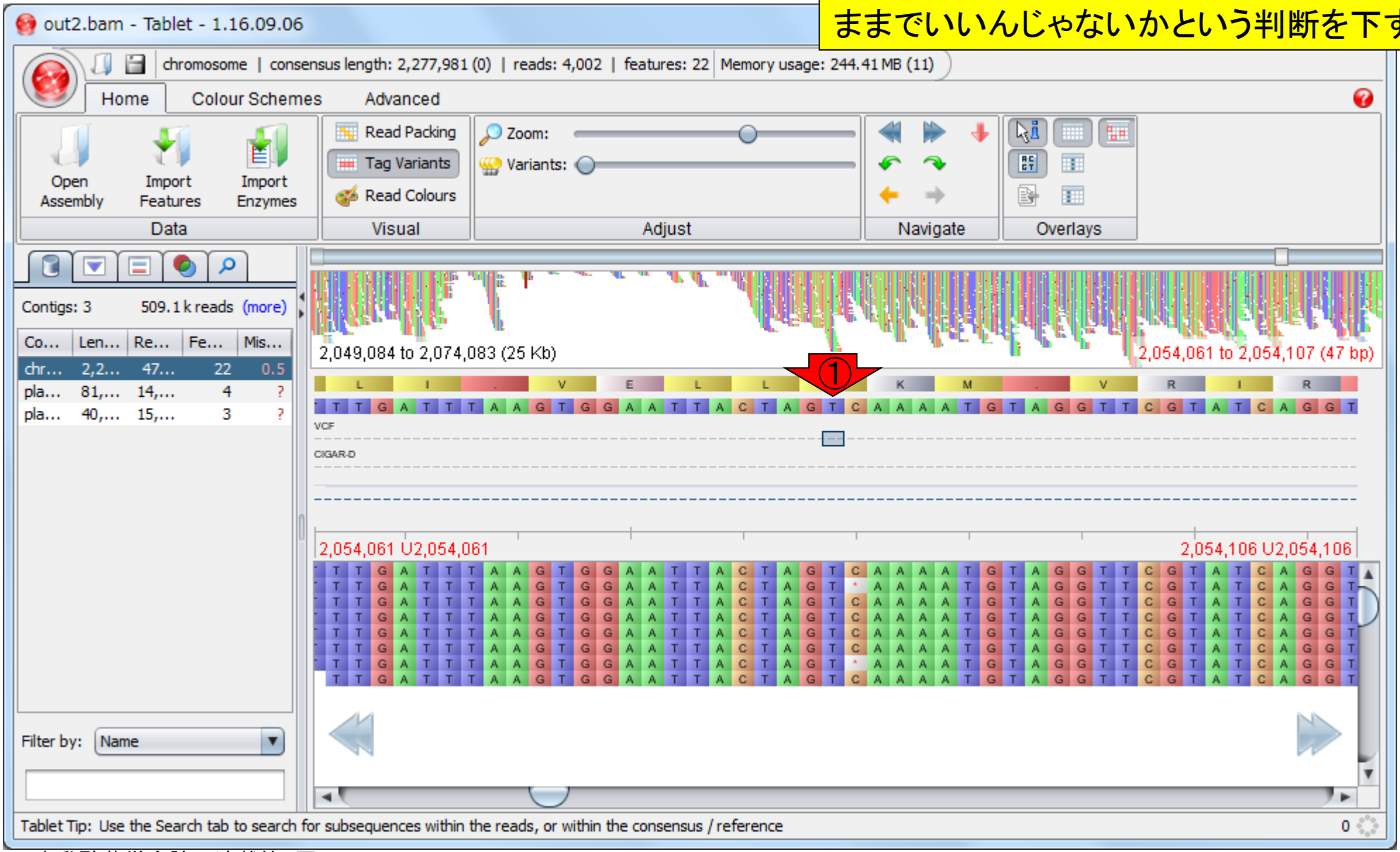
W22-2：2,054,084番目

①が指定した塩基ポジションにジャンプできるボタンなので押す。②2054084と入力して、③Padded Jump



W22-2：2,054,084番目

①2,054,084番目に移動完了。ここは「リファレンスはTCだけど、Tだけでいいんじゃないでしょうか」と提案されているところ。TCのままでいいんじゃないかという判断を下す



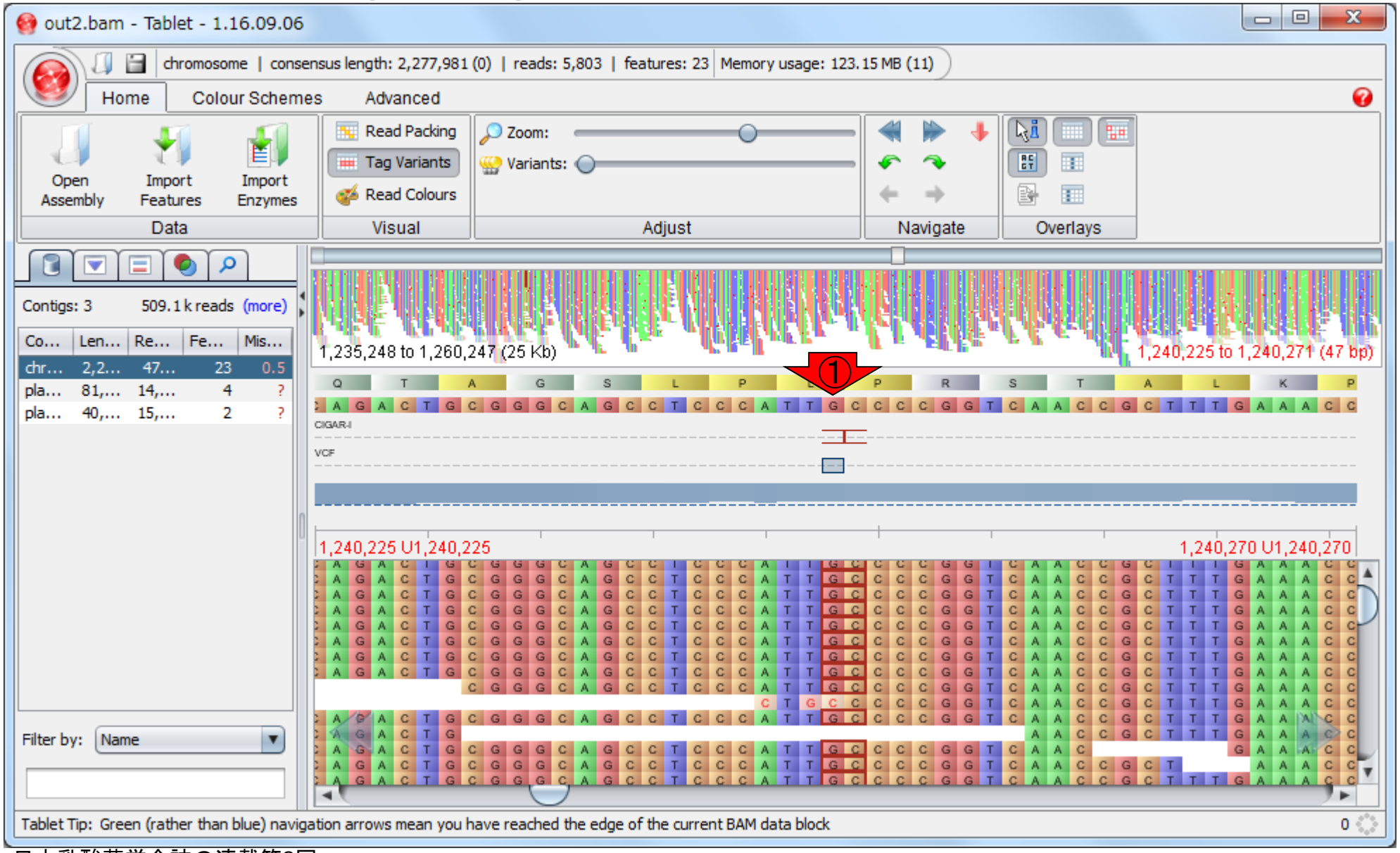
W22-3：1,240,248番目

次は、①1,240,248番目。これはリファレンス配列にCを1個挿入したほうがよいのでは、というもの

	A	B	C	D	E	F	G	H	I	J
27	#CHROM	POS	ID	REF	ALT	QUAL	FILTER	INFO	FORMAT	/home/w3pipeline/refdata/
28	chromosome	141598	.	A	G	17.1	.	DP=7;VDB=0.0077;AF1=0.5;AC1	GT:PL:GQ	0/1:47,0,138:50
29	chromosome	242819	.	T	C	20	.	DP=9;VDB=0.0111;AF1=0.5;AC1	GT:PL:GQ	0/1:50,0,139:53
30	chromosome	359141	.	T	C	4.13	.	DP=14;VDB=0.0037;AF1=0.4998	GT:PL:GQ	0/1:32,0,187:31
31	chromosome	463882	.	T	G	6.2	.	DP=13;VDB=0.0074;AF1=0.4999	GT:PL:GQ	0/1:35,0,216:36
32	chromosome	660031	.	C	T	8.64	.	DP=11;VDB=0.0063;AF1=0.5;AC	GT:PL:GQ	0/1:38,0,197:40
33	chromosome	792963	.	T	C	4.13	.	DP=14;VDB=0.0153;AF1=0.4998	GT:PL:GQ	0/1:32,0,200:31
34	chromosome	887422	.	G	A	10.4	.	DP=11;VDB=0.0070;AF1=0.5;AC	GT:PL:GQ	0/1:40,0,188:42
35	chromosome	913470	.	T	C	14.2	.	DP=9;VDB=0.0056;AF1=0.5;AC1	GT:PL:GQ	0/1:44,0,144:47
36	chromosome	965171	.	T	C	17.1	.	DP=20;VDB=0.0071;AF1=0.5;AC	GT:PL:GQ	0/1:47,0,249:50
37	chromosome	1134950	.	A	G	7.8	.	DP=23;VDB=0.0076;AF1=0.5;AC	GT:PL:GQ	0/1:37,0,255:39
38	chromosome	1160048	.	A	G	29	.	DP=17;VDB=0.0054;AF1=0.5;AC	GT:PL:GQ	0/1:59,0,255:62
39	chromosom	1160095	.	A	G	19.1	.	DP=10;VDB=0.0033;AF1=0.5;AC	GT:PL:GQ	0/1:49,0,191:52
40	chromoso	1240248	.	GCCCC	GCCCCC	214	.	INDEL;DP=48;VDB=0.0137;AF1=	GT:PL:GQ	1/1:255,128,0:99
41	chromosom	1455552	.	CAAAA	CAAAAA	214	.	INDEL;DP=58;VDB=0.0164;AF1=	GT:PL:GQ	1/1:255,130,0:99
42	chromosome	1475721	.	T	C	52	.	DP=19;VDB=0.0021;AF1=0.5;AC	GT:PL:GQ	0/1:82,0,246:85
43	chromosome	1752680	.	G	A	29	.	DP=24;VDB=0.0129;AF1=0.5;AC	GT:PL:GQ	0/1:59,0,255:62
44	chromosome	1941380	.	T	C	7.8	.	DP=13;VDB=0.0046;AF1=0.5;AC	GT:PL:GQ	0/1:37,0,214:39
45	chromosome	2052955	.	A	G	12.3	.	DP=8;VDB=0.0084;AF1=0.5;AC1	GT:PL:GQ	0/1:42,0,103:44
46	chromosome	2054084	.	TC	T	13.7	.	INDEL;DP=8;VDB=0.0009;AF1=0	GT:PL:GQ	0/1:51,0,174:54
47	chromosome	2056279	.	A	G	24	.	DP=6;VDB=0.0057;AF1=0.5;AC1	GT:PL:GQ	0/1:54,0,113:57
48	chromosome	2057564	.	T	C	8.64	.	DP=10;VDB=0.0041;AF1=0.5;AC	GT:PL:GQ	0/1:38,0,129:40
49	chromosome	2058145	.	C	T,G	108	.	DP=14;VDB=0.0147;AF1=1;AC1=	GT:PL:GQ	1/1:141,39,0,141,28,138:75
50	plasmid1	37177	.	G	T	54	.	DP=49;VDB=0.0087;AF1=0.5;AC	GT:PL:GQ	0/1:84,0,255:87
51	plasmid1	66609	.	A	G	225	.	DP=55;VDB=0.0060;AF1=0.5;AC	GT:PL:GQ	0/1:255,0,255:99
52	plasmid1	66637	.	C	T	225	.	DP=71;VDB=0.0141;AF1=0.5;AC	GT:PL:GQ	0/1:255,0,249:99
53	plasmid1	66668	.	C	T	225	.	DP=80;VDB=0.0153;AF1=0.5;AC	GT:PL:GQ	0/1:255,0,255:99
54	plasmid2	2569	.	CAAAAA	CAAAA	177	.	INDEL;DP=110;VDB=0.0166;AF1	GT:PL:GQ	1/1:218,255,0:99
55	plasmid2	37805	.	TAAAAA	TAAAA	104	.	INDEL;DP=31;VDB=0.0058;AF1=	GT:PL:GQ	1/1:145,93,0:99

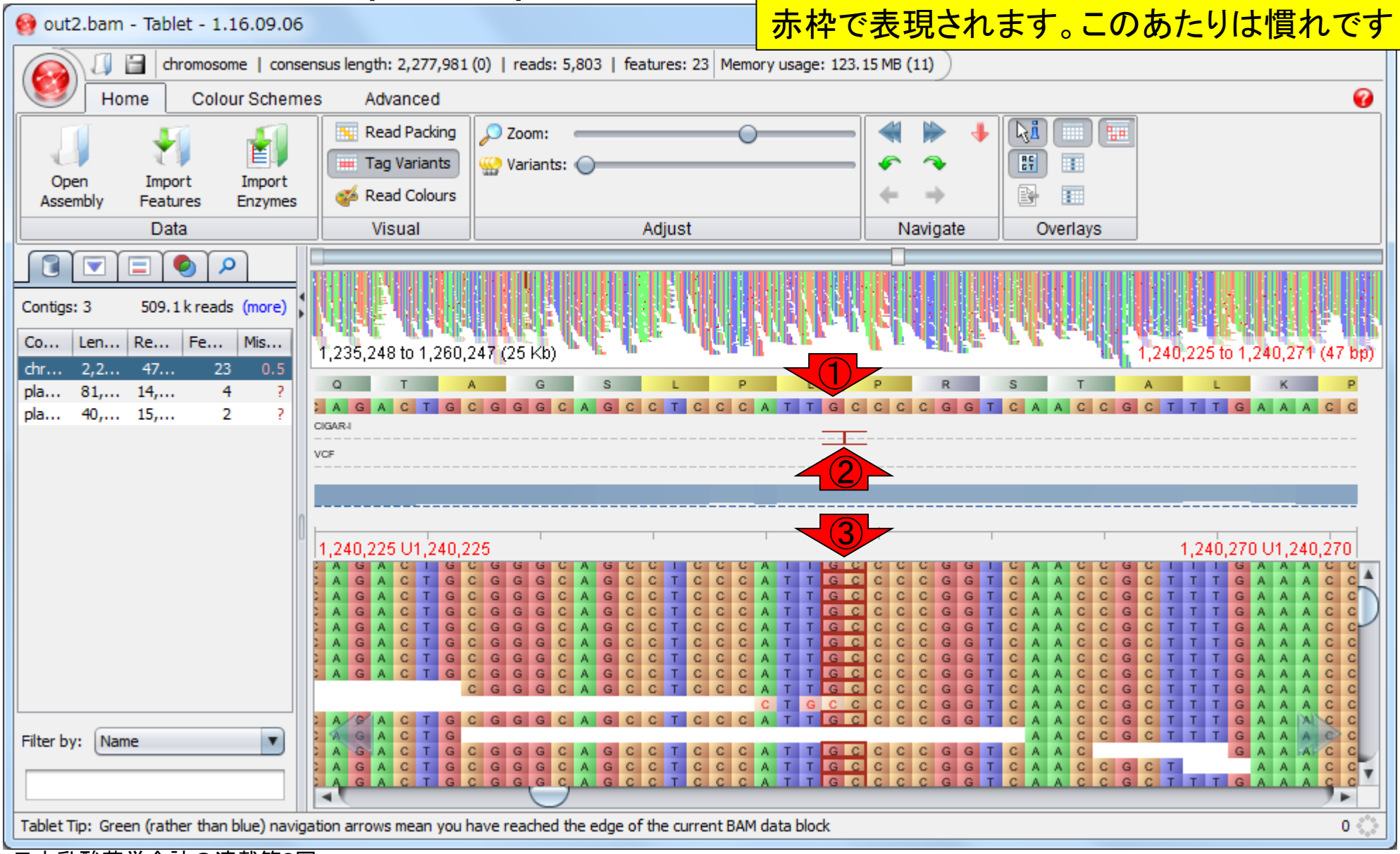
W22-3：1,240,248番目

こんな感じに見えます。①
指定したポジションはここ



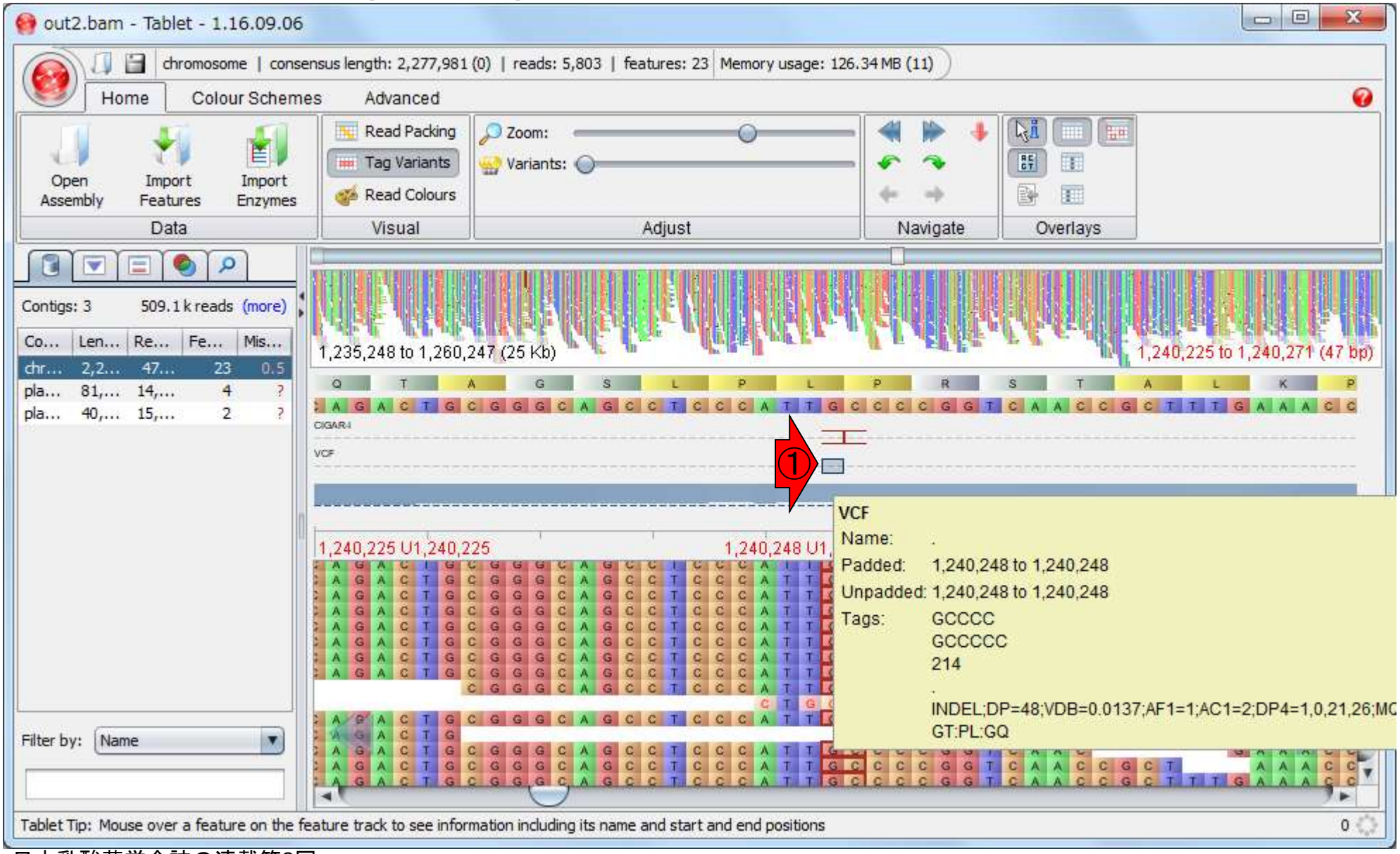
W22-3：1,240,248番目

②リファレンス配列のGとCの間に、insertion（挿入）の変異が入っている場合には、こんな感じに見えます。③マップされる側（Illuminaリード）は赤枠で表現されます。このあたりは慣れです



W22-3：1,240,248番目

①のあたりにカーソル移動させると、
Cのinsertionであることがわかります



次は、①2,058,145番目

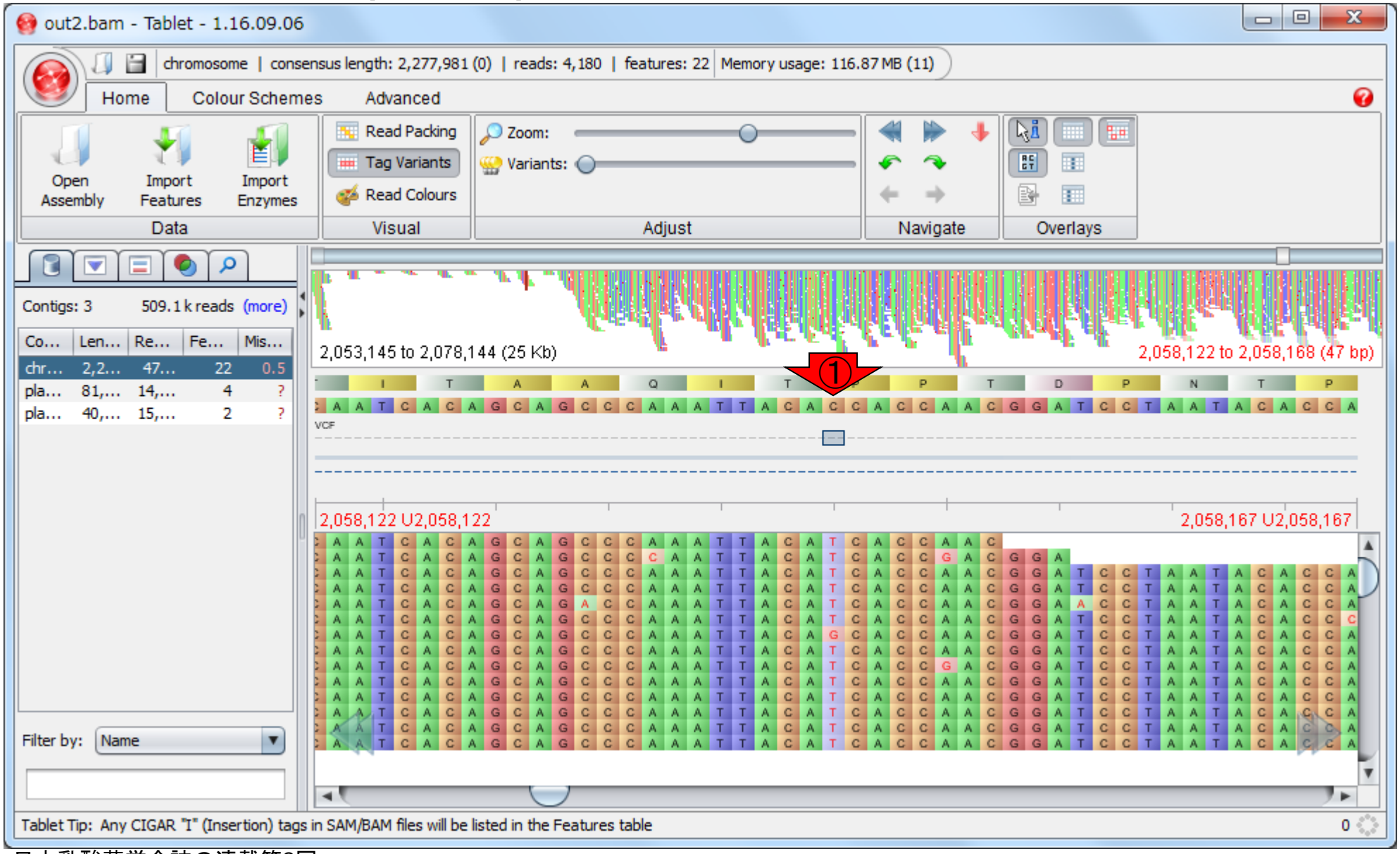
W22-4：2,058,145番目

	A	B	C	D	E	F	G	H	I	J
27	#CHROM	POS	ID	REF	ALT	QUAL	FILTER	INFO	FORMAT	/home/w3pipeline/refdata/
28	chromosome	141598	.	A	G	17.1	.	DP=7;VDB=0.0077;AF1=0.5;AC1	GT:PL:GQ	0/1:47,0,138:50
29	chromosome	242819	.	T	C	20	.	DP=9;VDB=0.0111;AF1=0.5;AC1	GT:PL:GQ	0/1:50,0,139:53
30	chromosome	359141	.	T	C	4.13	.	DP=14;VDB=0.0037;AF1=0.4998	GT:PL:GQ	0/1:32,0,187:31
31	chromosome	463882	.	T	G	6.2	.	DP=13;VDB=0.0074;AF1=0.4999	GT:PL:GQ	0/1:35,0,216:36
32	chromosome	660031	.	C	T	8.64	.	DP=11;VDB=0.0063;AF1=0.5;AC	GT:PL:GQ	0/1:38,0,197:40
33	chromosome	792963	.	T	C	4.13	.	DP=14;VDB=0.0153;AF1=0.4998	GT:PL:GQ	0/1:32,0,200:31
34	chromosome	887422	.	G	A	10.4	.	DP=11;VDB=0.0070;AF1=0.5;AC	GT:PL:GQ	0/1:40,0,188:42
35	chromosome	913470	.	T	C	14.2	.	DP=9;VDB=0.0056;AF1=0.5;AC1	GT:PL:GQ	0/1:44,0,144:47
36	chromosome	965171	.	T	C	17.1	.	DP=20;VDB=0.0071;AF1=0.5;AC	GT:PL:GQ	0/1:47,0,249:50
37	chromosome	1134950	.	A	G	7.8	.	DP=23;VDB=0.0076;AF1=0.5;AC	GT:PL:GQ	0/1:37,0,255:39
38	chromosome	1160048	.	A	G	29	.	DP=17;VDB=0.0054;AF1=0.5;AC	GT:PL:GQ	0/1:59,0,255:62
39	chromosome	1160095	.	A	G	19.1	.	DP=10;VDB=0.0033;AF1=0.5;AC	GT:PL:GQ	0/1:49,0,191:52
40	chromosome	1240248	.	GCCCC	GCCCCC	214	.	INDEL;DP=48;VDB=0.0137;AF1=	GT:PL:GQ	1/1:255,128,0:99
41	chromosome	1455552	.	CAAAA	CAAAAA	214	.	INDEL;DP=58;VDB=0.0164;AF1=	GT:PL:GQ	1/1:255,130,0:99
42	chromosome	1475721	.	T	C	52	.	DP=19;VDB=0.0021;AF1=0.5;AC	GT:PL:GQ	0/1:82,0,246:85
43	chromosome	1752680	.	G	A	29	.	DP=24;VDB=0.0129;AF1=0.5;AC	GT:PL:GQ	0/1:59,0,255:62
44	chromosome	1941380	.	T	C	7.8	.	DP=13;VDB=0.0046;AF1=0.5;AC	GT:PL:GQ	0/1:37,0,214:39
45	chromosome	2052955	.	A	G	12.3	.	DP=8;VDB=0.0084;AF1=0.5;AC1	GT:PL:GQ	0/1:42,0,103:44
46	chromosome	2054084	.	TC	T	13.7	.	INDEL;DP=8;VDB=0.0009;AF1=0	GT:PL:GQ	0/1:51,0,174:54
47	chromosome	2056279	.	A	G	24	.	DP=6;VDB=0.0057;AF1=0.5;AC1	GT:PL:GQ	0/1:54,0,113:57
48	chromosom	2057564	.	T	C	8.64	.	DP=10;VDB=0.0041;AF1=0.5;AC	GT:PL:GQ	0/1:38,0,129:40
49	chromosom	2058145	.	C	T,G	108	.	DP=14;VDB=0.0147;AF1=1;AC1=	GT:PL:GQ	1/1:141,39,0,141,28,138:75
50	plasmid1	37177	.	G	T	54	.	DP=49;VDB=0.0087;AF1=0.5;AC	GT:PL:GQ	0/1:84,0,255:87
51	plasmid1	66609	.	A	G	225	.	DP=55;VDB=0.0060;AF1=0.5;AC	GT:PL:GQ	0/1:255,0,255:99
52	plasmid1	66637	.	C	T	225	.	DP=71;VDB=0.0141;AF1=0.5;AC	GT:PL:GQ	0/1:255,0,249:99
53	plasmid1	66668	.	C	T	225	.	DP=80;VDB=0.0153;AF1=0.5;AC	GT:PL:GQ	0/1:255,0,255:99
54	plasmid2	2569	.	CAAAAA	CAAAA	177	.	INDEL;DP=110;VDB=0.0166;AF1	GT:PL:GQ	1/1:218,255,0:99
55	plasmid2	37805	.	TAAAAA	TAAAA	104	.	INDEL;DP=31;VDB=0.0058;AF1=	GT:PL:GQ	1/1:145,93,0:99

①

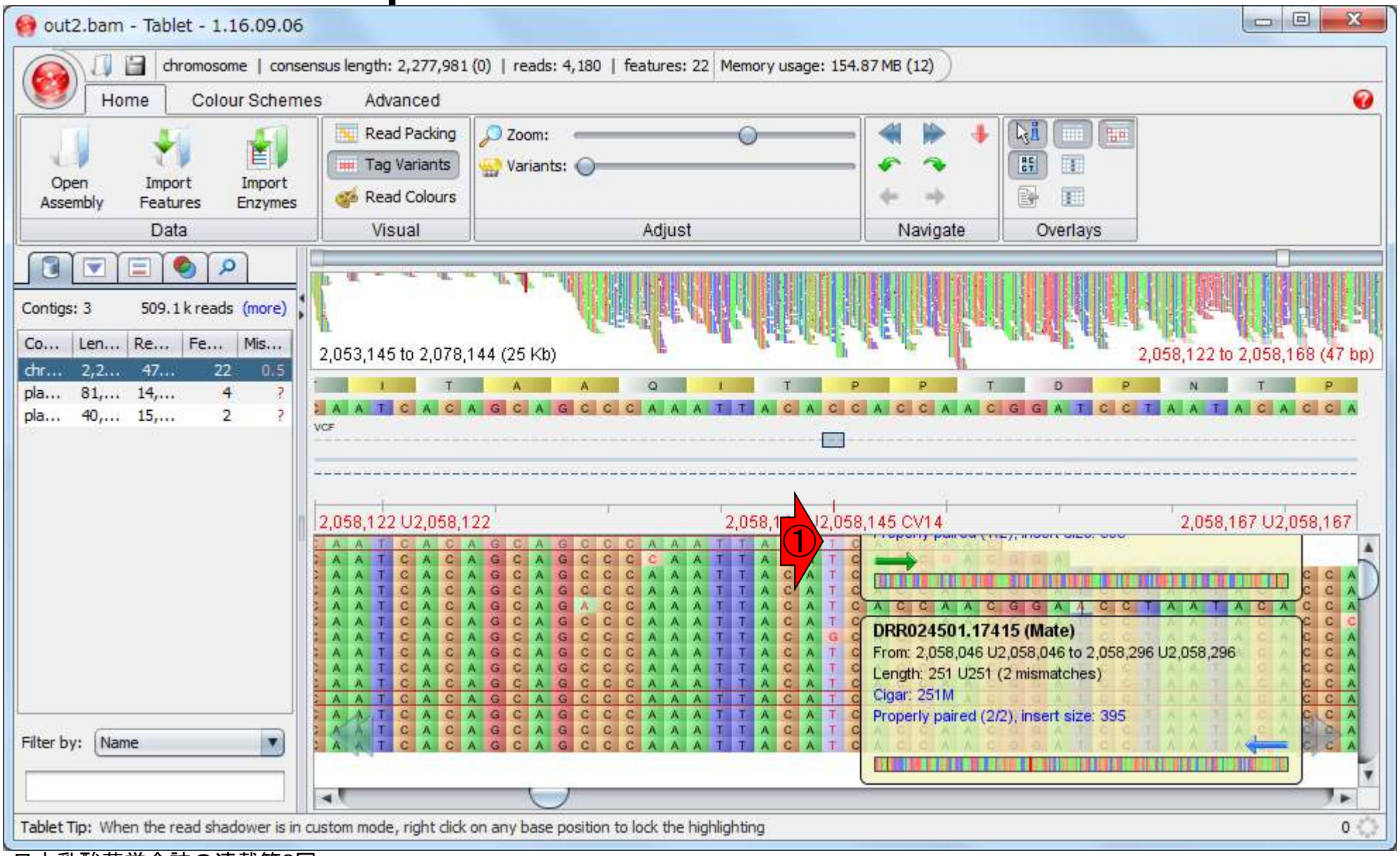
W22-4：2,058,145番目

①リファレンス中のCではなく、TやGとなっていたが、これはTに置換したほうがよいと判断



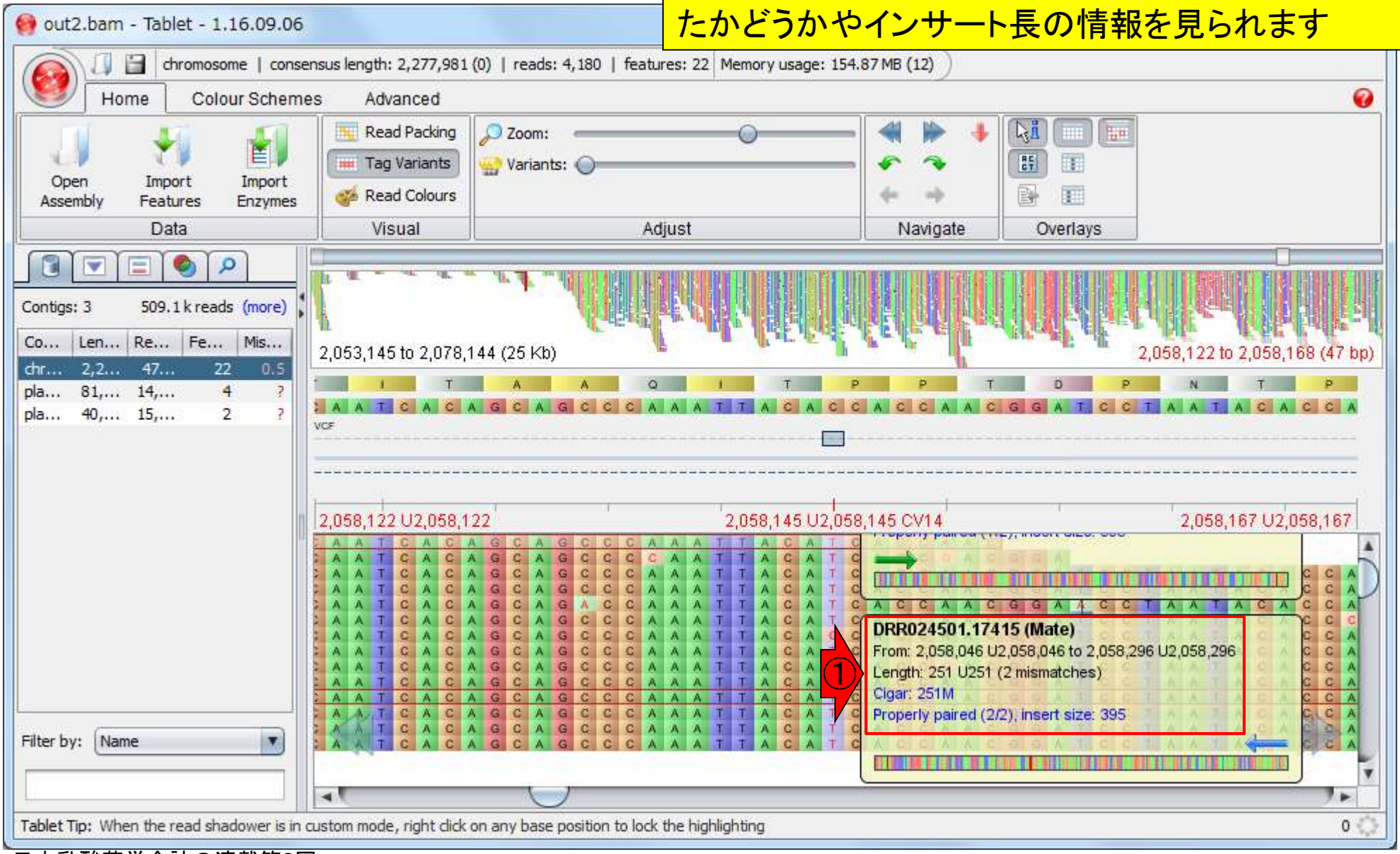
W23-1：Tips（リード情報）

①のあたりにポインタを移動させたところ。リードごとの情報を見ることができます



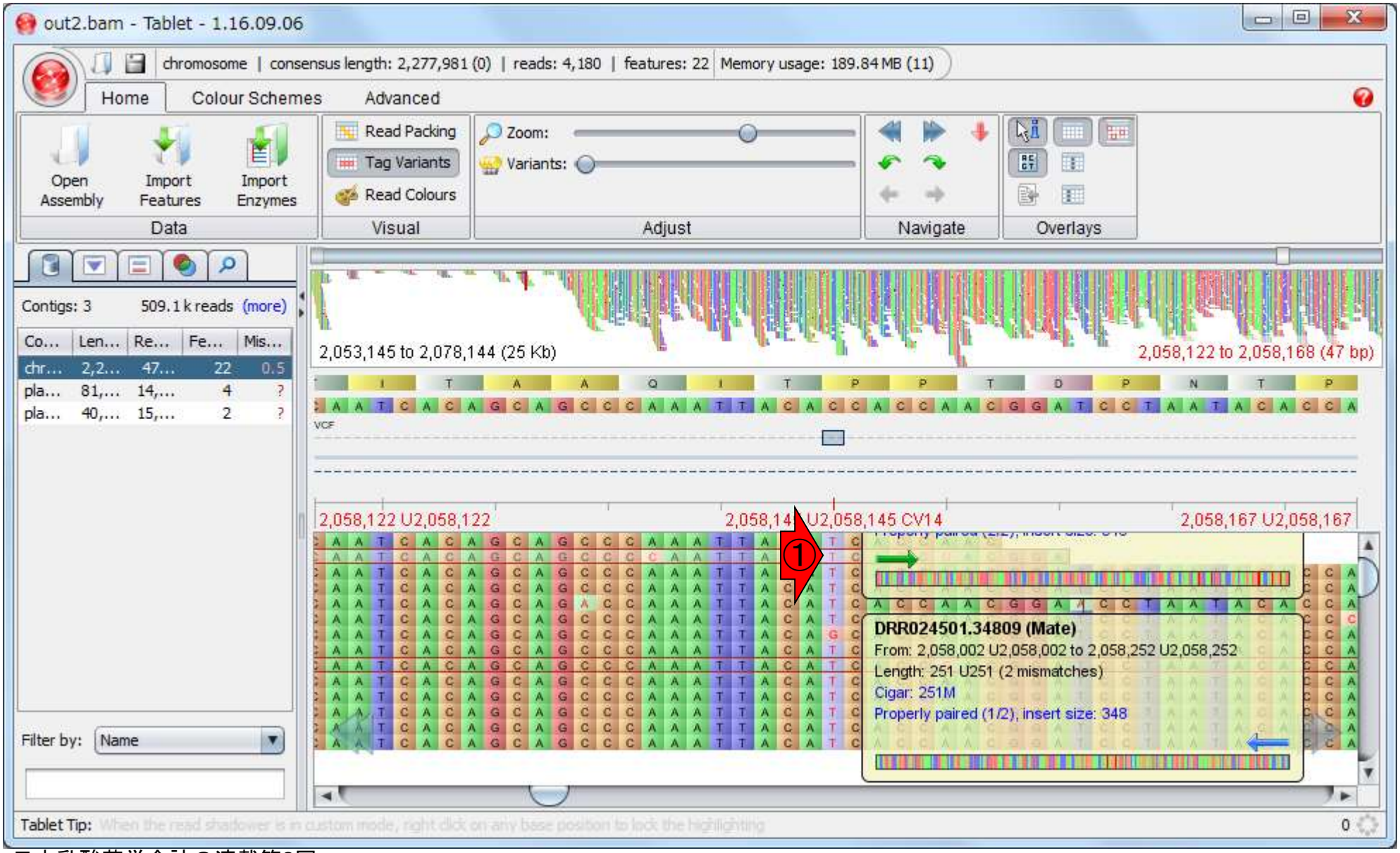
W23-1：リード情報

①赤枠の上から順に、(1) リードID、(2)マップされた領域、(3)長さとミスマッチ数、(4)CIGAR string、(5)paired-endでマップしているのでもう片方（メイト）もマップされたかどうかやインサート長の情報を見られます



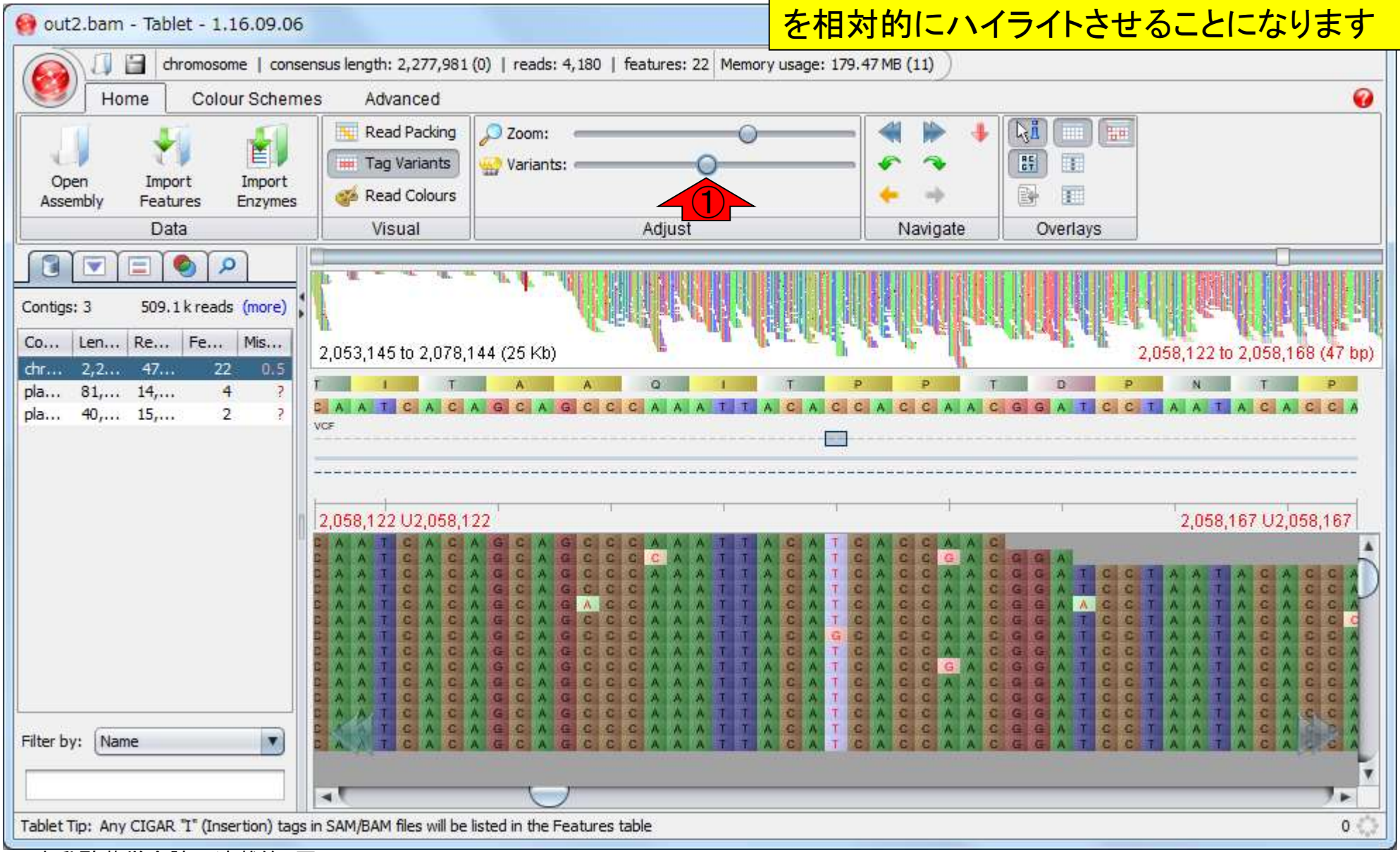
W23-1：リード情報

この位置（リファレンス配列の2,058,145番目の塩基上）にマップされた2つめのリードはこんな感じ



W23-2：Variants

①Variantsのところを右に移動させることは、リファレンス配列と異なる塩基(variants)以外を暗くさせていくことに相当します。つまりvariantsを相対的にハイライトさせることになります



①右端まで移動させるとこんな感じ

W23-2：Variants

全体を眺めて下した方針。灰色で示した
計5か所が概要配列に変更を加えるところ

W24-1：方針のまとめ

#CHROM	POS	REF	ALT	方針	理由
chromosome	141598	A	G	Aのまま	Aが5個、Gが2個だったから
chromosome	242819	T	C	Tのまま	Tが7個、Cが2個だったから
chromosome	359141	T	C	Tのまま	Tが多数派だから
chromosome	463882	T	G	Tのまま	Tが2個、Gが1個だったから
chromosome	660031	C	T	Cのまま	Cが8個、Tが2個だったから
chromosome	792963	T	C	Tのまま	Tが多数派だから
chromosome	887422	G	A	Gのまま	Gが多数派だから
chromosome	913470	T	C	Tのまま	Tが多数派だから
chromosome	965171	T	C	Tのまま	Tが多数派だから
chromosome	1134950	A	G	Aのまま	Aが多数派だから
chromosome	1160048	A	G	Aのまま	Aが多数派だから
chromosome	1160095	A	G	Aのまま	Aが多数派だから
chromosome	1240248	GCCCC	GCCCCC	Cを挿入	そのほうが多数派だから
chromosome	1455552	CAAAA	CAAAAA	Aを挿入	そのほうが多数派だから
chromosome	1475721	T	C	Tのまま	Tが多数派だから
chromosome	1752680	G	A	Gのまま	Gが多数派だから
chromosome	1941380	T	C	Tのまま	Tが多数派だから
chromosome	2052955	A	G	Aのまま	Aが多数派だから
chromosome	2054084	TC	T	TCのまま	Cがあるほうが多数派だから
chromosome	2056279	A	G	Aのまま	Aが多数派だから
chromosome	2057564	T	C	Tのまま	Tが多数派だから
chromosome	2058145	C	T,G	Tに変更	Tが多数派だから
plasmid1	37177	G	T	Gのまま	Gが多数派だから
plasmid1	66609	A	G	Aのまま	Gのほうがちょっと多いが微妙
plasmid1	66637	C	T	Cのまま	Tのほうがちょっと多いが微妙
plasmid1	66668	C	T	Cのまま	Tのほうがちょっと多いが微妙
plasmid2	2569	CAAAAA	CAAAA	Aを欠失	そのほうが多数派だから
plasmid2	37805	TAAAAA	TAAAA	Aを欠失	そのほうが多数派だから

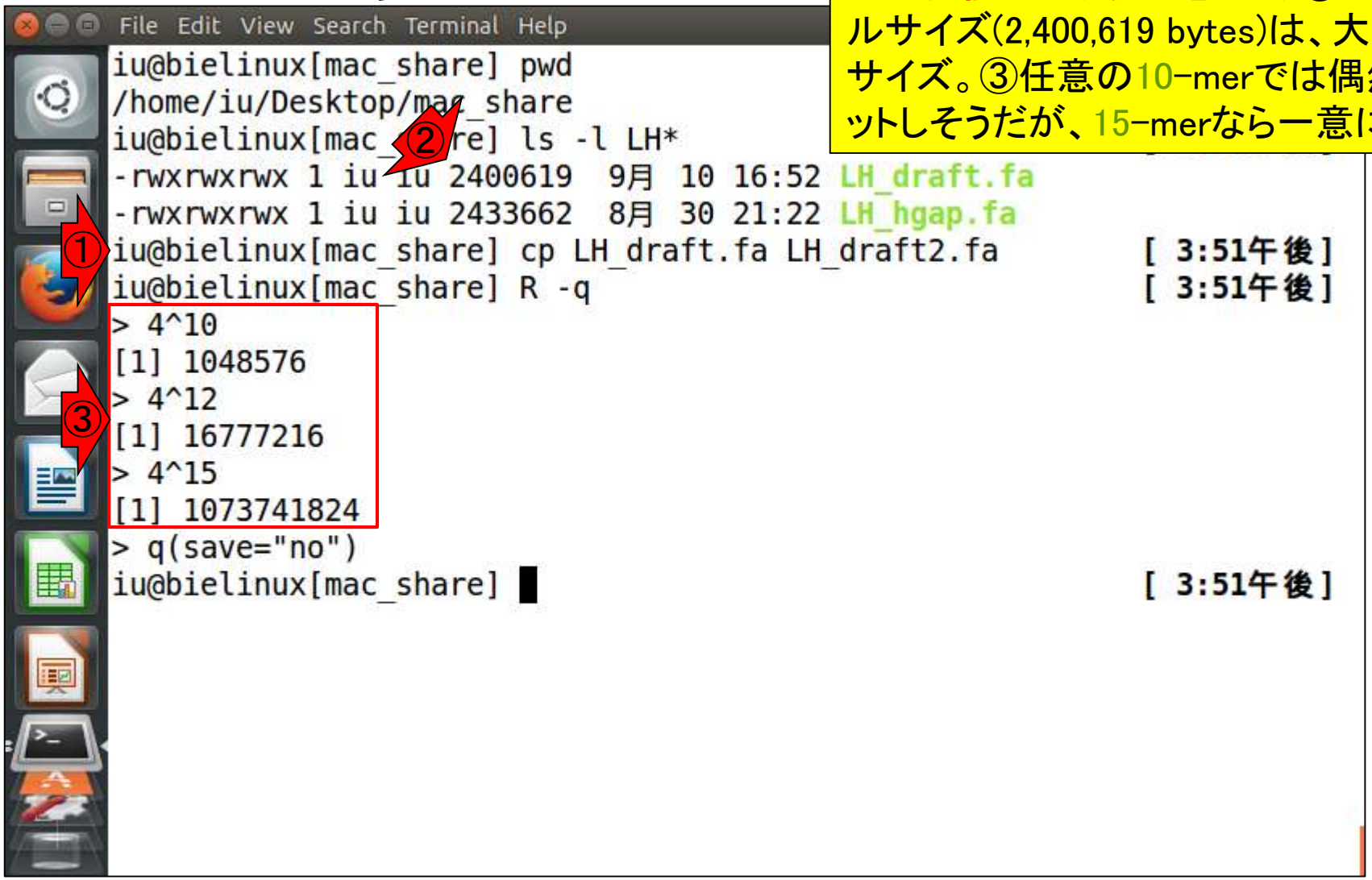
W24-1：VCFと一致

当たり前と言えば当たり前ですが、VCFファイルから読み取れる結論とほぼ同じになります(W18-4)

	A	B	C	D	E	F	G	H	I	J
27	#CHROM	POS	ID	REF	ALT	QUAL	FILTER	INFO	FORMAT	/home/w3pipeline/refdata/
28	chromosome	141598	.	A	G	17.1	.	DP=7;VDB=0.0077;AF1=0.5;AC1	GT:PL:GQ	0/1:47,0,138:50
29	chromosome	242819	.	T	C	20	.	DP=9;VDB=0.0111;AF1=0.5;AC1	GT:PL:GQ	0/1:50,0,139:53
30	chromosome	359141	.	T	C	4.13	.	DP=14;VDB=0.0037;AF1=0.4998	GT:PL:GQ	0/1:32,0,187:31
31	chromosome	463882	.	T	G	6.2	.	DP=13;VDB=0.0074;AF1=0.4999	GT:PL:GQ	0/1:35,0,216:36
32	chromosome	660031	.	C	T	8.64	.	DP=11;VDB=0.0063;AF1=0.5;AC	GT:PL:GQ	0/1:38,0,197:40
33	chromosome	792963	.	T	C	4.13	.	DP=14;VDB=0.0153;AF1=0.4998	GT:PL:GQ	0/1:32,0,200:31
34	chromosome	887422	.	G	A	10.4	.	DP=11;VDB=0.0070;AF1=0.5;AC	GT:PL:GQ	0/1:40,0,188:42
35	chromosome	913470	.	T	C	14.2	.	DP=9;VDB=0.0056;AF1=0.5;AC1	GT:PL:GQ	0/1:44,0,144:47
36	chromosome	965171	.	T	C	17.1	.	DP=20;VDB=0.0071;AF1=0.5;AC	GT:PL:GQ	0/1:47,0,249:50
37	chromosome	1134950	.	A	G	7.8	.	DP=23;VDB=0.0076;AF1=0.5;AC	GT:PL:GQ	0/1:37,0,255:39
38	chromosome	1160048	.	A	G	29	.	DP=17;VDB=0.0054;AF1=0.5;AC	GT:PL:GQ	0/1:59,0,255:62
39	chromosome	1160095	.	A	G	19.1	.	DP=10;VDB=0.0033;AF1=0.5;AC	GT:PL:GQ	0/1:49,0,191:52
40	chromosome	1240248	.	GCCCC	GCCCCC	214	.	INDEL;DP=48;VDB=0.0137;AF1=	GT:PL:GQ	1/1:255,128,0:99
41	chromosome	1455552	.	CAAAA	CAAAAA	214	.	INDEL;DP=58;VDB=0.0164;AF1=	GT:PL:GQ	1/1:255,130,0:99
42	chromosome	1475721	.	T	C	52	.	DP=19;VDB=0.0021;AF1=0.5;AC	GT:PL:GQ	0/1:82,0,246:85
43	chromosome	1752680	.	G	A	29	.	DP=24;VDB=0.0129;AF1=0.5;AC	GT:PL:GQ	0/1:59,0,255:62
44	chromosome	1941380	.	T	C	7.8	.	DP=13;VDB=0.0046;AF1=0.5;AC	GT:PL:GQ	0/1:37,0,214:39
45	chromosome	2052955	.	A	G	12.3	.	DP=8;VDB=0.0084;AF1=0.5;AC1	GT:PL:GQ	0/1:42,0,103:44
46	chromosome	2054084	.	TC	T	13.7	.	INDEL;DP=8;VDB=0.0009;AF1=0	GT:PL:GQ	0/1:51,0,174:54
47	chromosome	2056279	.	A	G	24	.	DP=6;VDB=0.0057;AF1=0.5;AC1	GT:PL:GQ	0/1:54,0,113:57
48	chromosome	2057564	.	T	C	8.64	.	DP=10;VDB=0.0041;AF1=0.5;AC	GT:PL:GQ	0/1:38,0,129:40
49	chromosome	2058145	.	C	T,G	108	.	DP=14;VDB=0.0147;AF1=1;AC1=	GT:PL:GQ	1/1:41,39,0,141,28,138:75
50	plasmid1	37177	.	G	T	54	.	DP=49;VDB=0.0087;AF1=0.5;AC	GT:PL:GQ	0/1:84,0,255:87
51	plasmid1	66609	.	A	G	225	.	DP=55;VDB=0.0060;AF1=0.5;AC	GT:PL:GQ	0/1:255,0,255:99
52	plasmid1	66637	.	C	T	225	.	DP=71;VDB=0.0141;AF1=0.5;AC	GT:PL:GQ	0/1:255,0,249:99
53	plasmid1	66668	.	C	T	225	.	DP=80;VDB=0.0153;AF1=0.5;AC	GT:PL:GQ	0/1:255,0,255:99
54	plasmid2	2569	.	CAAAAA	CAAAA	177	.	INDEL;DP=110;VDB=0.0166;AF1	GT:PL:GQ	1/1:218,255,0:99
55	plasmid2	37805	.	TAAAAA	TAAAA	104	.	INDEL;DP=31;VDB=0.0058;AF1=	GT:PL:GQ	1/1:45,93,0:99

W26-1：変異の反映

①変異反映後のファイルをLH_draft2.faにすべく、LH_draft.faからコピー。ここでは、変異の反映を文字列置換で行う戦略をとる。②オリジナルのファイルサイズ(2,400,619 bytes)は、大まかにはゲノムサイズ。③任意の10-merでは偶然に複数個所ヒットしそうだが、15-merなら一意に決まりそうだ



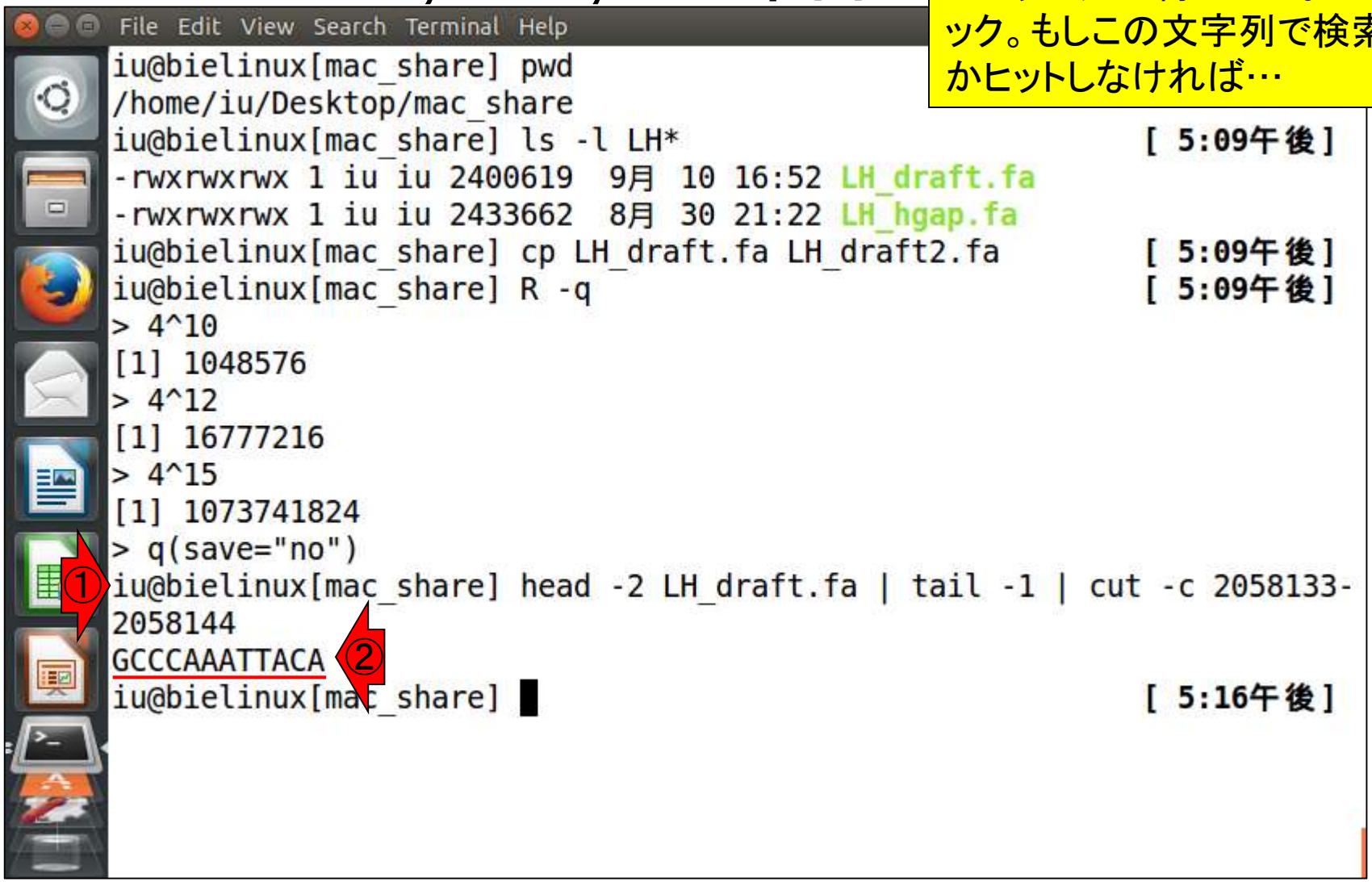
W26-1：2,058,145番目

①まずはchromosome上の2,058,145番目のCをTに置換すべく、上流12塩基分（つまり2058133-2058144）の塩基配列を抽出。配列ごとに位置を表す数値が大きいものから順番に行えば、位置のずれに悩まされることはない。だから一番最後からやろうとしている

#CHROM	POS	REF	ALT	方針	
chromosome	141598	A	G	Aのまま	Aが5個、
chromosome	242819	T	C	Tのまま	Tが7個、
chromosome	359141	T	C	Tのまま	T
chromosome	463882	T	G	Tのまま	Tが2個、Gが1個だったから
chromosome	660031	C	T	Cのまま	Cが8個、Tが2個だったから
chromosome	792963	T	C	Tのまま	Tが多数派だから
chromosome	887422	G	A	Gのまま	Gが多数派だから
chromosome	913470	T	C	Tのまま	Tが多数派だから
chromosome	965171	T	C	Tのまま	Tが多数派だから
chromosome	1134950	A	G	Aのまま	Aが多数派だから
chromosome	1160048	A	G	Aのまま	Aが多数派だから
chromosome	1160095	A	G	Aのまま	Aが多数派だから
chromosome	1240248	GCCCC	GCCCCC	Cを挿入	そのほうが多数派だから
chromosome	1455552	CAAAA	CAAAAA	Aを挿入	そのほうが多数派だから
chromosome	1475721	T	C	Tのまま	Tが多数派だから
chromosome	1752680	G	A	Gのまま	Gが多数派だから
chromosome	1941380	T	C	Tのまま	Tが多数派だから
chromosome	2052955	A	G	Aのまま	Aが多数派だから
chromosome	2054084	TC	T	TCのまま	Cがあるほうが多数派だから
chromosome	2056279	A	G	Aのまま	Aが多数派だから
chromosome	2057564	T	C	Tのまま	Tが多数派だから
chromosome	2058145	C	T,G	Tに変更	Tが多数派だから
plasmid1	37177	G	T	Gのまま	Gが多数派だから
plasmid1	66609	A	G	Aのまま	Gのほうがちょっと多いが微妙
plasmid1	66637	C	T	Cのまま	Tのほうがちょっと多いが微妙
plasmid1	66668	C	T	Cのまま	Tのほうがちょっと多いが微妙
plasmid2	2569	CAAAAA	CAAAA	Aを欠失	そのほうが多数派だから
plasmid2	37805	TAAAAA	TAAAA	Aを欠失	そのほうが多数派だから

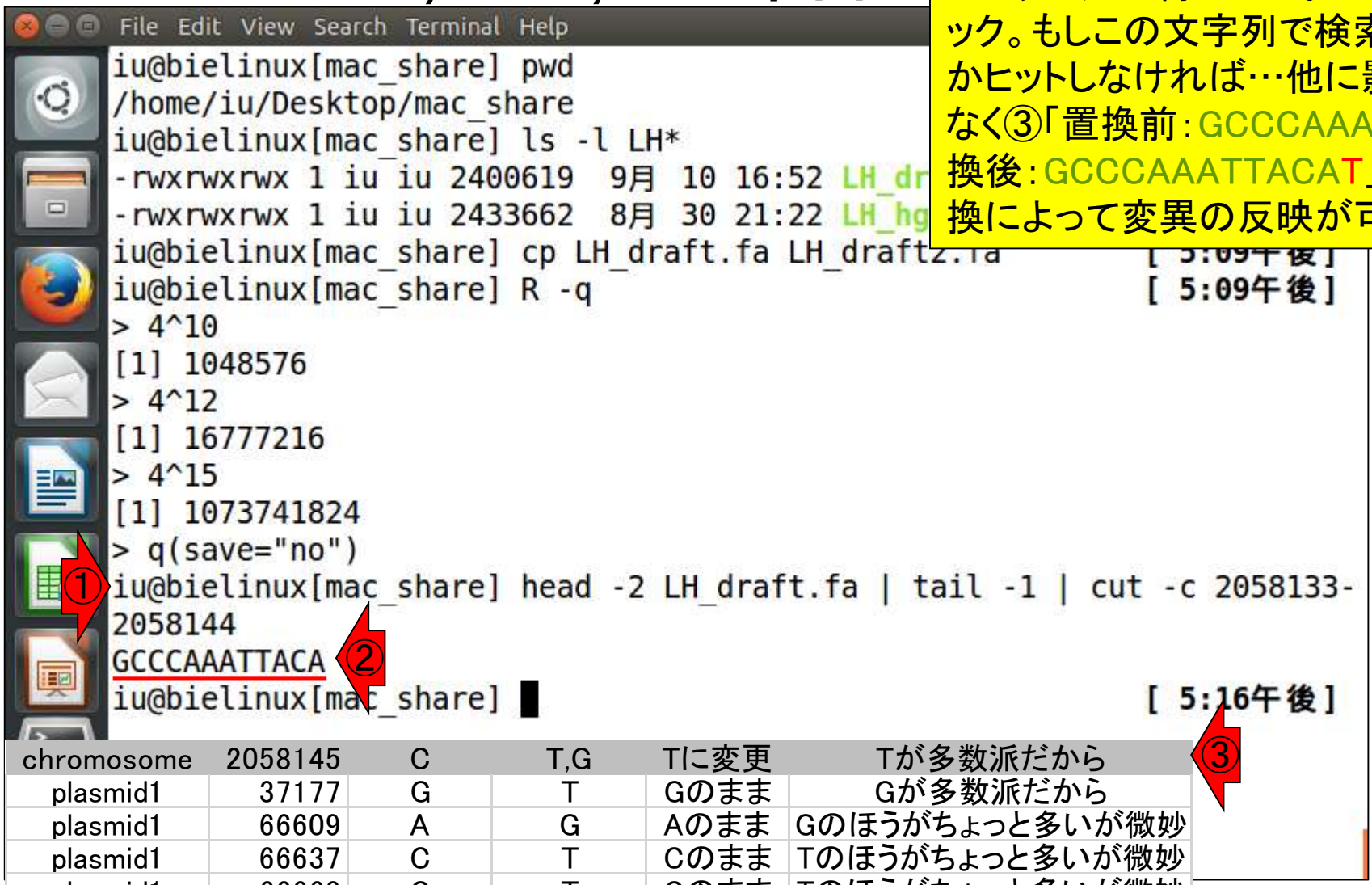
W26-1：2,058,145番目

①chromosome上の(2058133-2058144)番目の塩基配列を抽出。②GCCCAAATTACAという配列が得られた。W11-2と同じテクニック。もしこの文字列で検索して1箇所にしかヒットしなければ…



W26-1：2,058,145番目

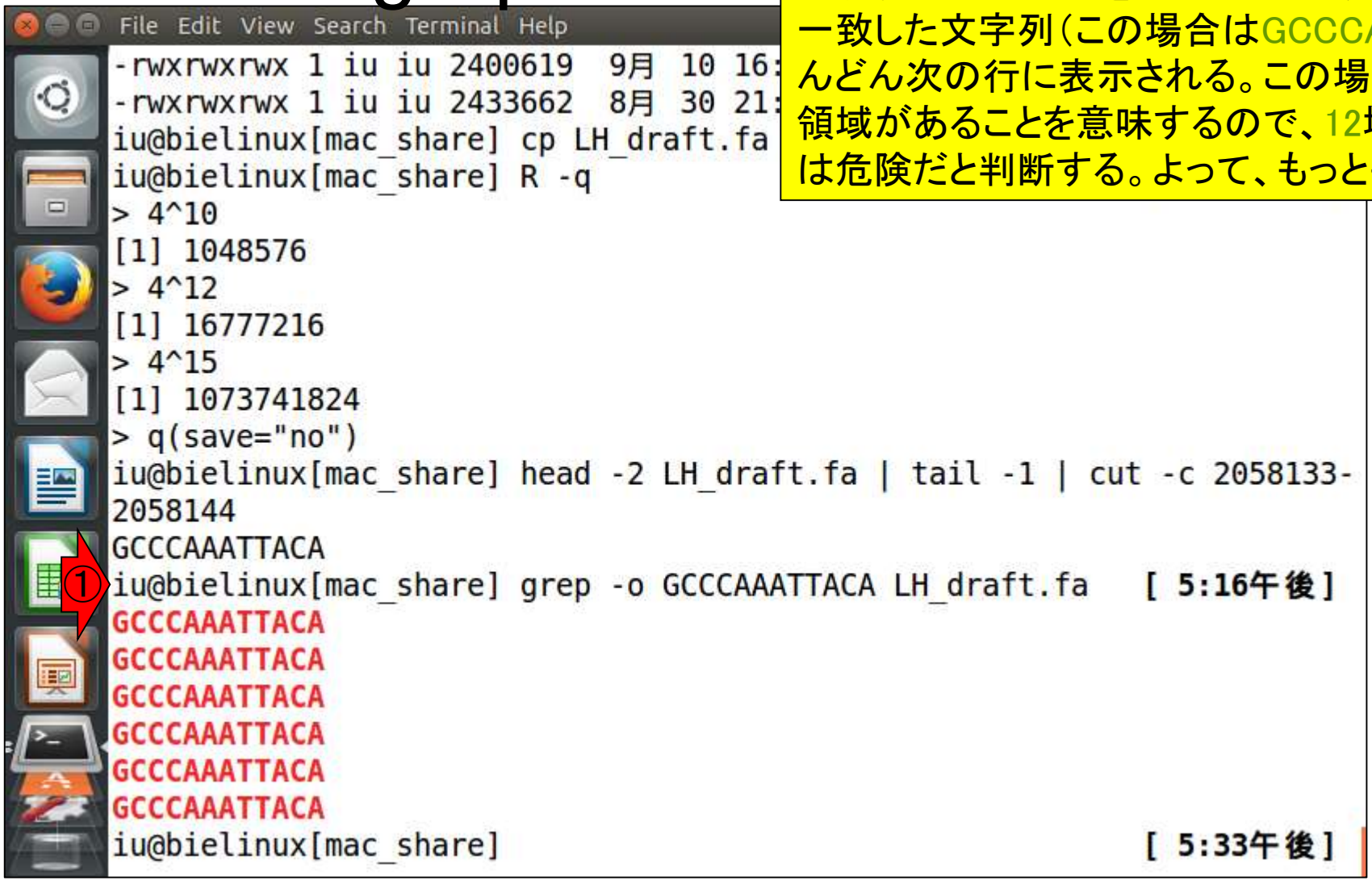
①chromosome上の(2058133-2058144)番目の塩基配列を抽出。②GCCCAAATTACAという配列が得られた。W11-2と同じテクニック。もしこの文字列で検索して1箇所にしかヒットしなければ…他に影響を及ぼすことなく③「置換前：GCCCAAATTACAC」、「置換後：GCCCAAATTACAT」という文字列置換によって変異の反映が可能（という戦略）



chromosome	2058145	C	T,G	Tに変更	Tが多数派だから
plasmid1	37177	G	T	Gのまま	Gが多数派だから
plasmid1	66609	A	G	Aのまま	Gのほうがちょっと多いが微妙
plasmid1	66637	C	T	Cのまま	Tのほうがちょっと多いが微妙
plasmid1	66668	C	T	Cのまま	Tのほうがちょっと多いが微妙
plasmid2	2569	CAAAAA	CAAAA	Aを欠失	そのほうが多数派だから
plasmid2	37805	TAAAAA	TAAAA	Aを欠失	そのほうが多数派だから

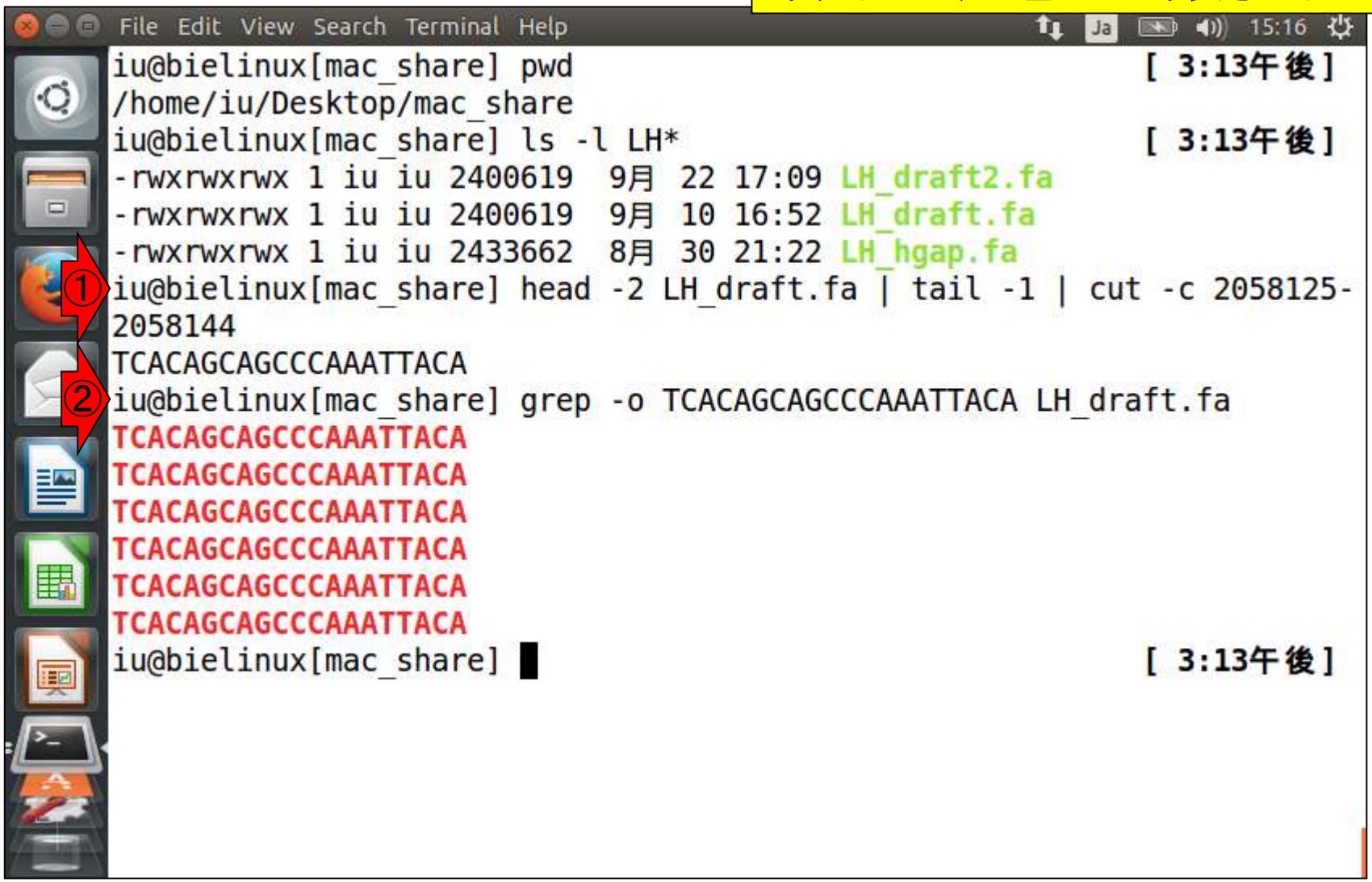
W26-2：grep -o

①grepのデフォルトはGCCCAAATTACAという文字列と一致した行全体を返すため、一致箇所の個数は不明。-oオプションをつけることで、一致箇所分だけ一致した文字列（この場合はGCCCAAATTACA）がどんどん次の行に表示される。この場合は6箇所一致領域があることを意味するので、12塩基という長さでは危険だと判断する。よって、もっと長い塩基にする



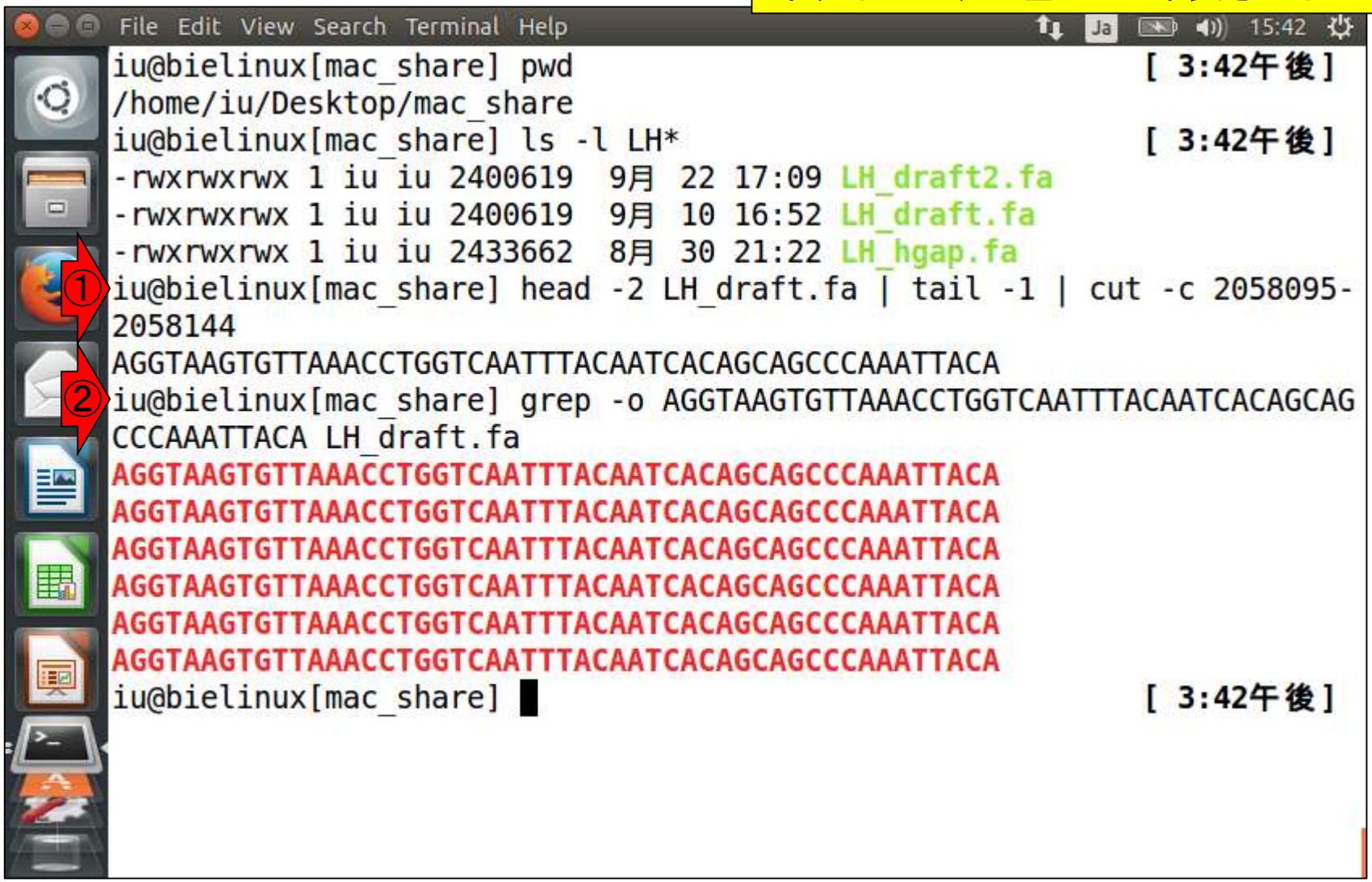
W26-3：20塩基で

①chromosome上の(2058125-2058144)番目の塩基配列を抽出。この場合でも6箇所一致領域があることを意味するので、20塩基という長さでもまだ危険だと判断



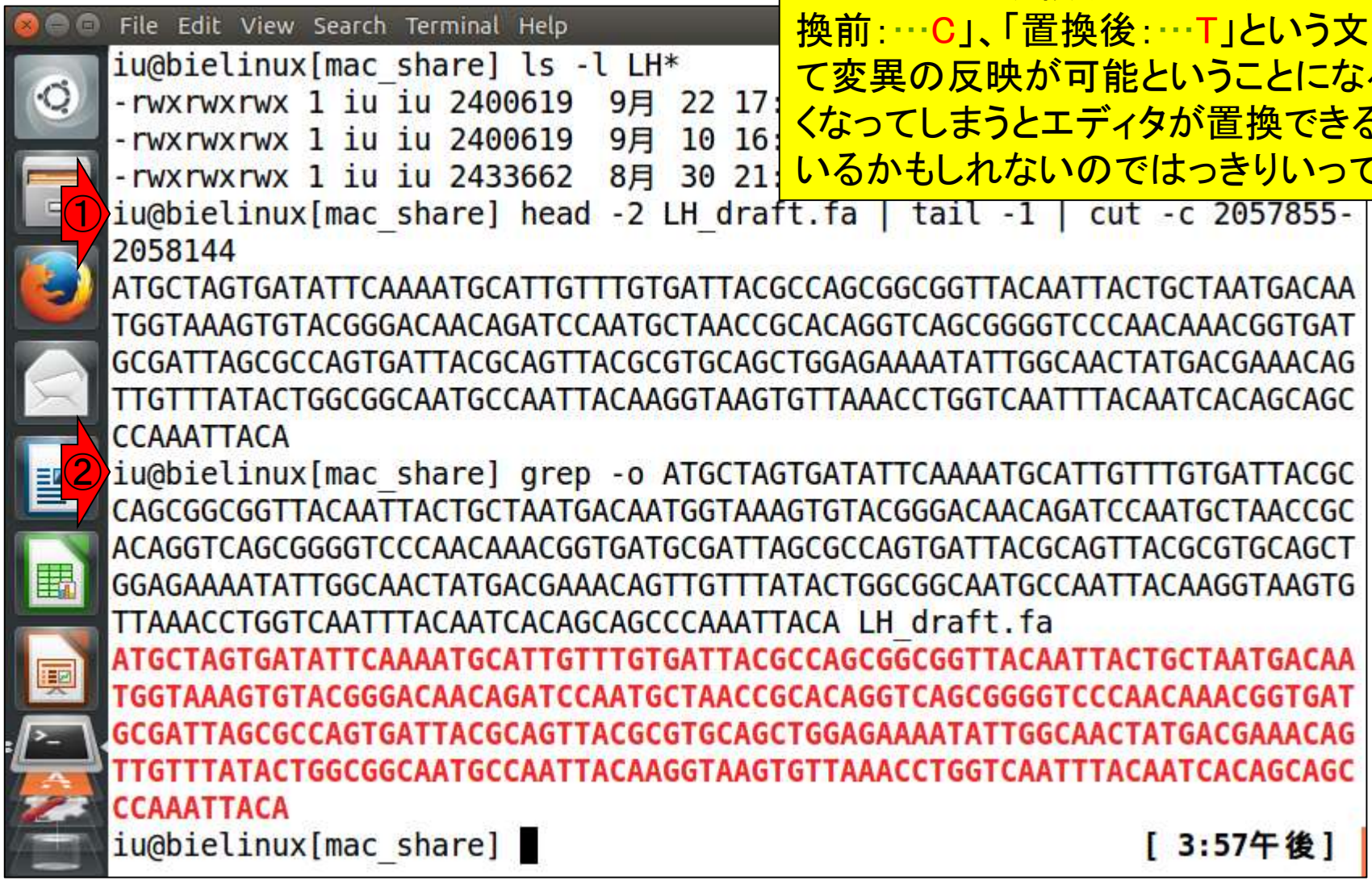
W26-4：50塩基で

①chromosome上の(2058095-2058144)番目の塩基配列を抽出。この場合でも6箇所一致領域があることを意味するので、50塩基という長さでもまだ危険だと判断



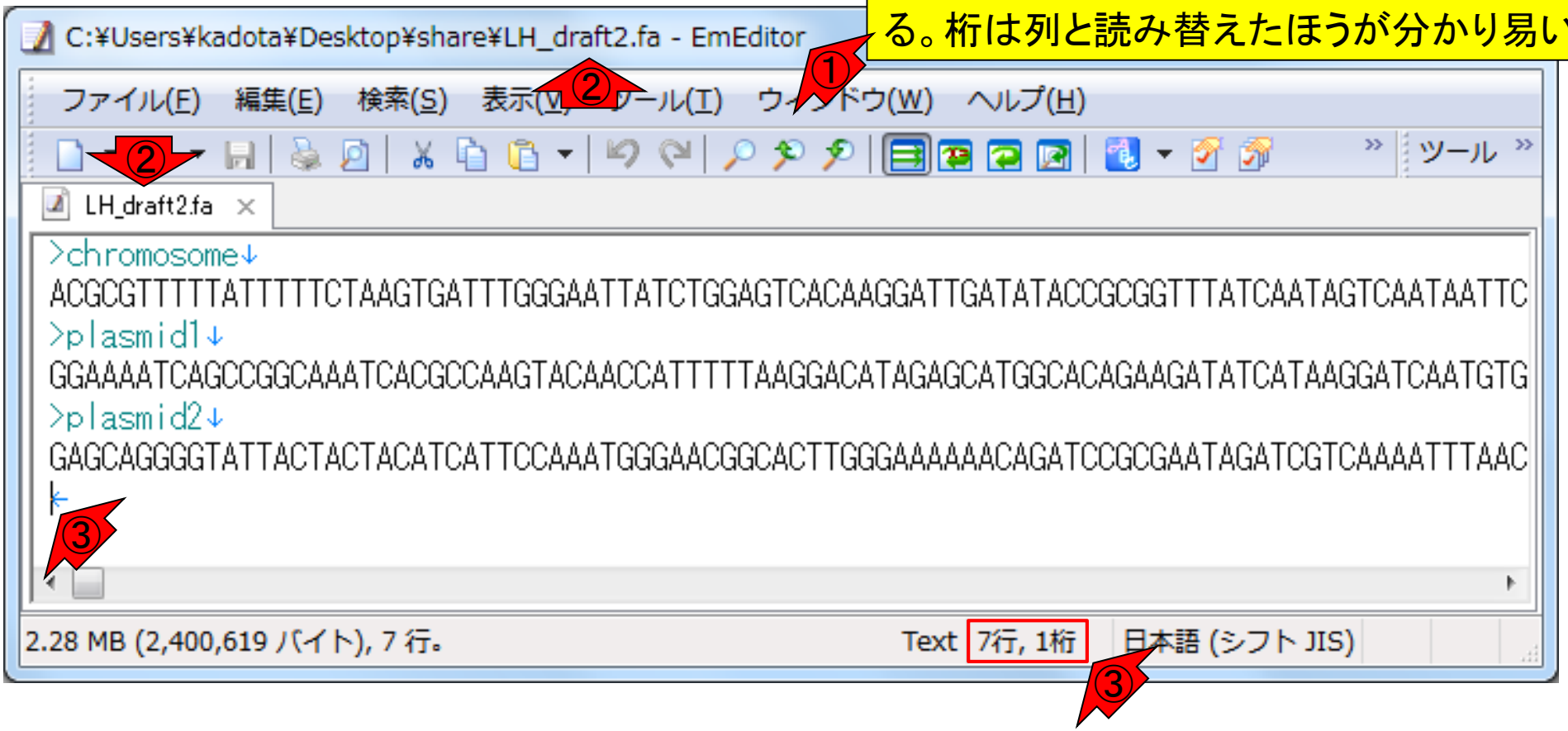
W26-5：300塩基で

①chromosome上の(2057855-2058144)番目の塩基配列を抽出。200塩基ではだめだったが300塩基の長さでやっと一致領域が一か所のみとなったので、「置換前：…C」、「置換後：…T」という文字列置換によって変異の反映が可能ということになるが、これだけ長くなってしまうとエディタが置換できる最大値を超えているかもしれないのではっきりいってアホラシイ



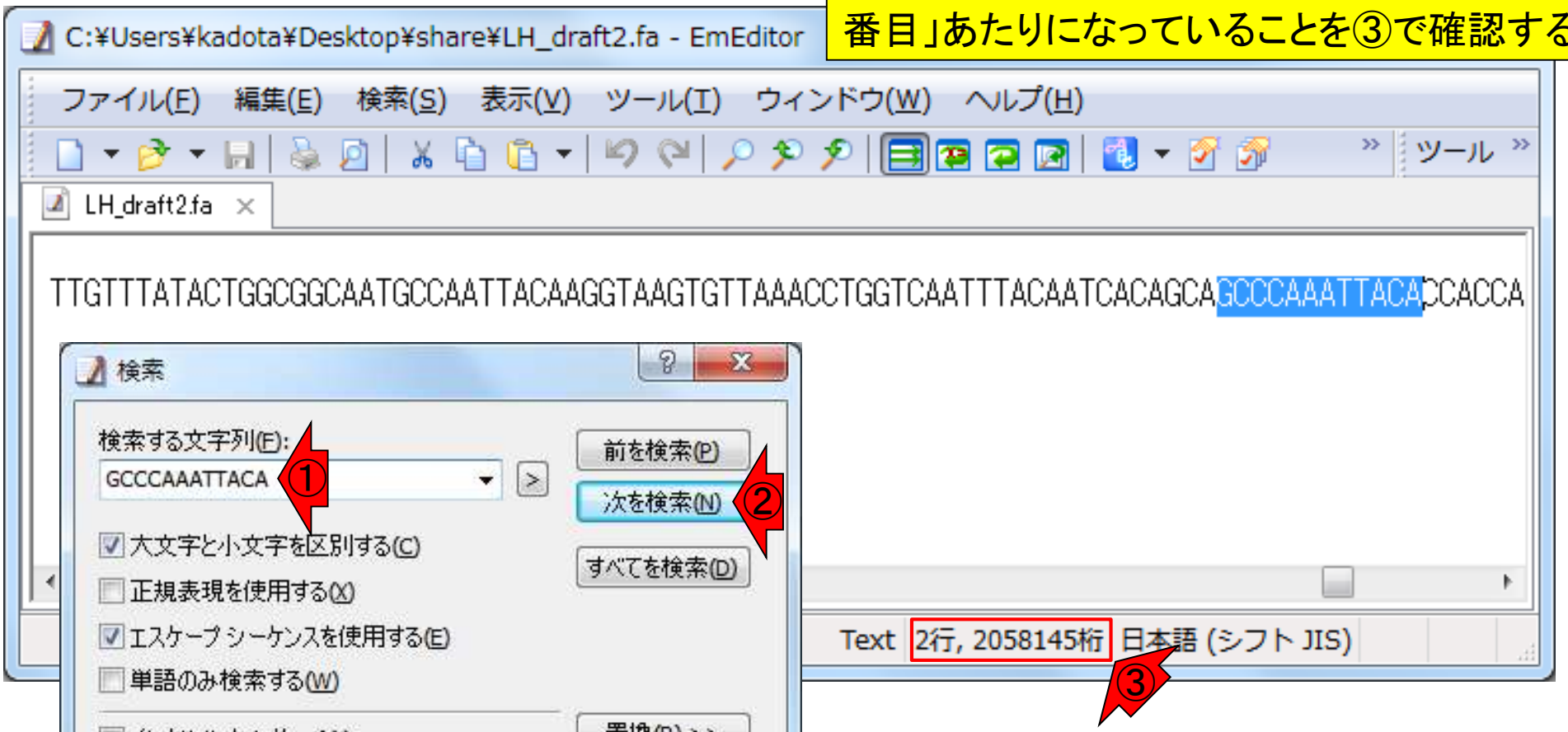
W27-1：EmEditor

①EmEditor（Windows用のみ）で、②変異箇所を変更する予定のLH_draft2.faを開いたところ。③現在のカーソル位置が7行1桁であることがわかる。桁は列と読み替えたほうが分かり易いかも…



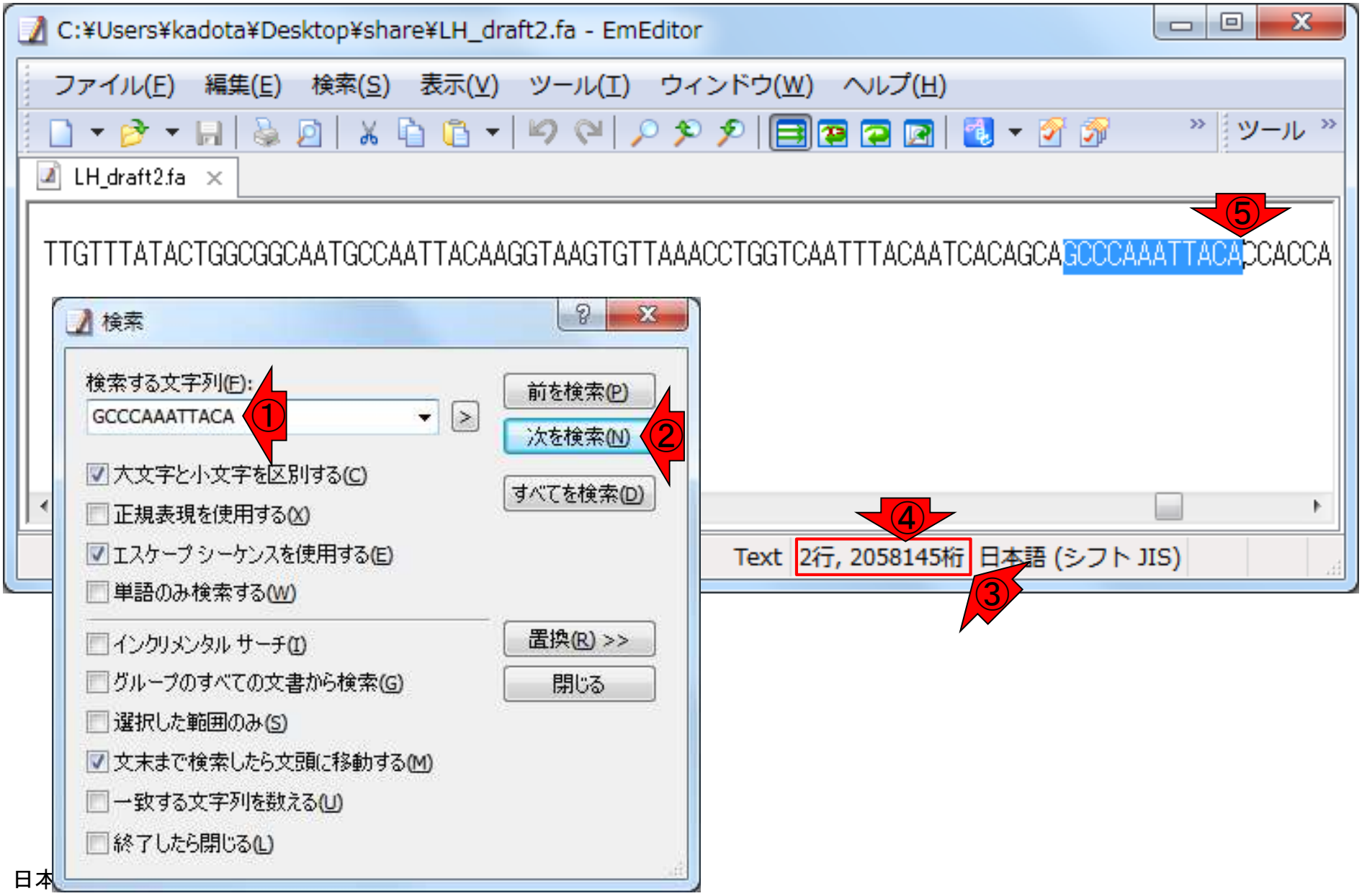
W27-2：EmEditor

①W26-2で6箇所の一致領域があることが分かっているGCCCAAATTACAという文字列で検索する。②を何回か押し、目的の「chromosome上の2,058,145番目」あたりになっていることを③で確認する



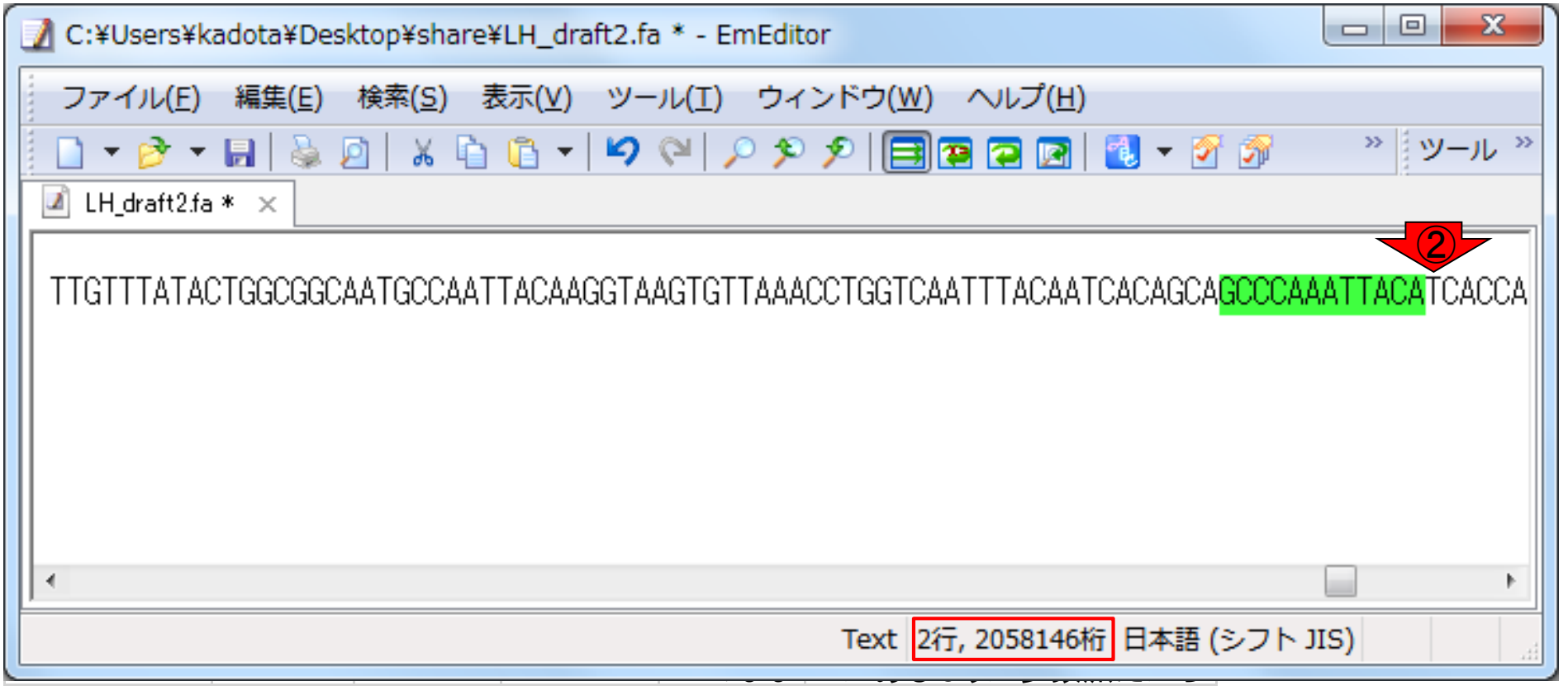
W27-2：EmEditor

④そのものズバリの位置になっているが、⑤の位置にカーソルがあるので妥当



W27-2：EmEditor

①chromosome上の2,058,145番目の、②CをTに変更したところ



chromosome	2056279	A	G	Aのまま	Aが多数派だから
chromosome	2057564	T	C	Tのまま	Tが多数派だから
chromosome	2058145	C	T,G	Tに変更	Tが多数派だから
plasmid1	37177	G	T	Gのまま	Gが多数派だから
plasmid1	66609	A	G	Aのまま	Gのほうがちょっと多いが微妙
plasmid1	66637	C	T	Cのまま	Tのほうがちょっと多いが微妙
plasmid1	66668	C	T	Cのまま	Tのほうがちょっと多いが微妙
plasmid2	2569	CAAAAA	CAAAA	Aを欠失	そのほうが多数派だから
plasmid2	37805	TAAAAA	TAAAA	Aを欠失	そのほうが多数派だから

①

次は欠失の反映例として①plasmid2上の37,805番目のところを示す

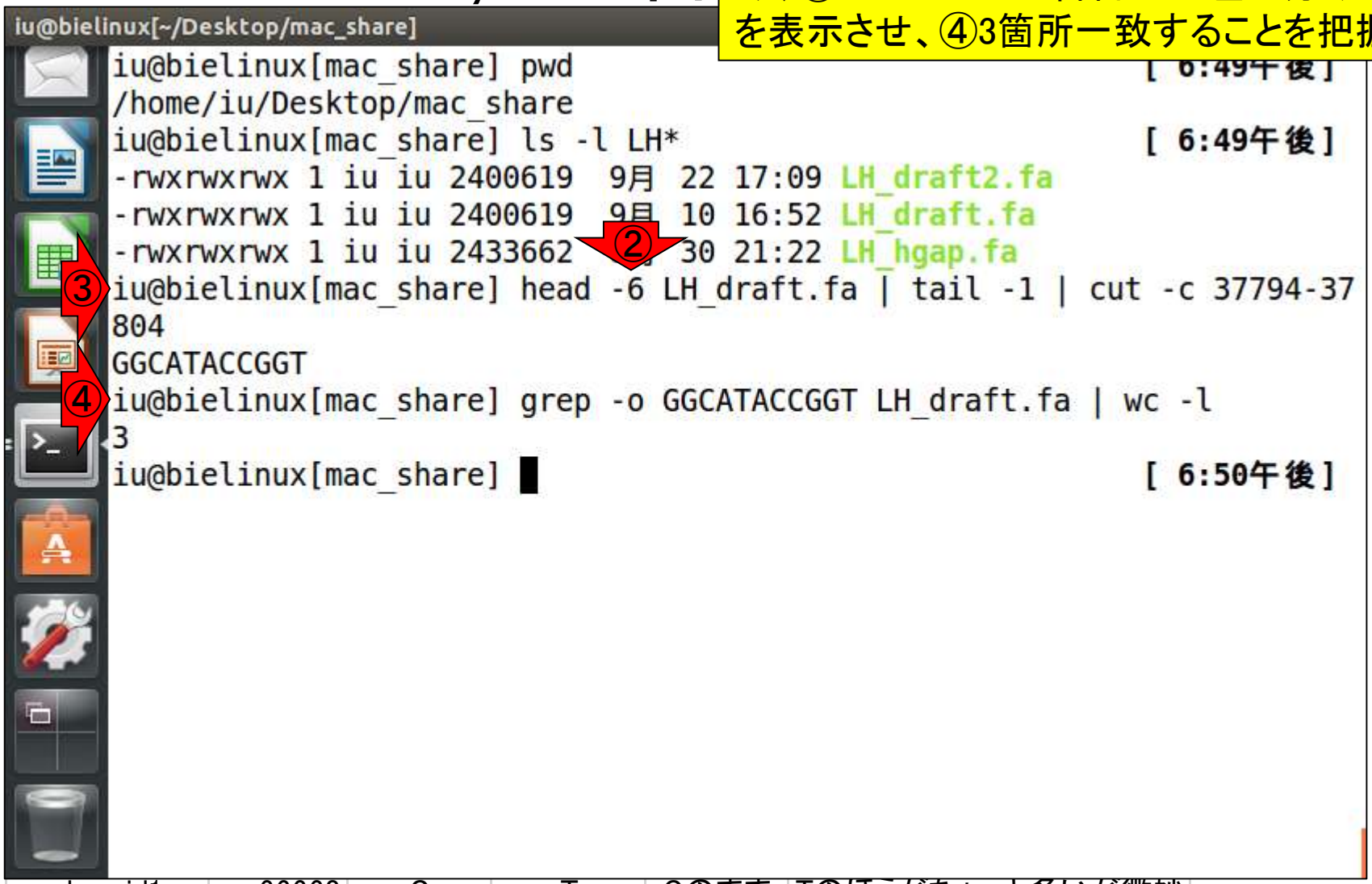
W27-3:37,805番目

#CHROM	POS	REF	ALT	方針	理由
chromosome	141598	A	G	Aのまま	Aが5個、Gが2個だったから
chromosome	242819	T	C	Tのまま	Tが7個、Cが2個だったから
chromosome	359141	T	C	Tのまま	Tが多数派だから
chromosome	463882	T	G	Tのまま	Tが2個、Gが1個だったから
chromosome	660031	C	T	Cのまま	Cが8個、Tが2個だったから
chromosome	792963	T	C	Tのまま	Tが多数派だから
chromosome	887422	G	A	Gのまま	Gが多数派だから
chromosome	913470	T	C	Tのまま	Tが多数派だから
chromosome	965171	T	C	Tのまま	Tが多数派だから
chromosome	1134950	A	G	Aのまま	Aが多数派だから
chromosome	1160048	A	G	Aのまま	Aが多数派だから
chromosome	1160095	A	G	Aのまま	Aが多数派だから
chromosome	1240248	GCCCC	GCCCCC	Cを挿入	そのほうが多数派だから
chromosome	1455552	CAAAA	CAAAAA	Aを挿入	そのほうが多数派だから
chromosome	1475721	T	C	Tのまま	Tが多数派だから
chromosome	1752680	G	A	Gのまま	Gが多数派だから
chromosome	1941380	T	C	Tのまま	Tが多数派だから
chromosome	2052955	A	G	Aのまま	Aが多数派だから
chromosome	2054084	TC	T	TCのまま	Cがあるほうが多数派だから
chromosome	2056279	A	G	Aのまま	Aが多数派だから
chromosome	2057564	T	C	Tのまま	Tが多数派だから
chromosome	2058145	C	T,G	Tに変更	Tが多数派だから
plasmid1	37177	G	T	Gのまま	Gが多数派だから
plasmid1	66609	A	G	Aのまま	Gのほうがちょっと多いが微妙
plasmid1	66637	C	T	Cのまま	Tのほうがちょっと多いが微妙
plasmid1	66668	C	T	Cのまま	Tのほうがちょっと多いが微妙
plasmid2	2569	CAAAAA	CAAAA	Aを欠失	そのほうが多数派だから
plasmid2	37805	TAAAAA	TAAAA	Aを欠失	そのほうが多数派だから

①

W27-3：37,805番

①plasmid2上の37,805番目を合理的に検索すべく、②plasmid2上の（FASTA形式ファイルで3番目の配列なので6）、③37794-37804番目の11塩基分（GGCATACCGGT）を表示させ、④3箇所一致することを把握

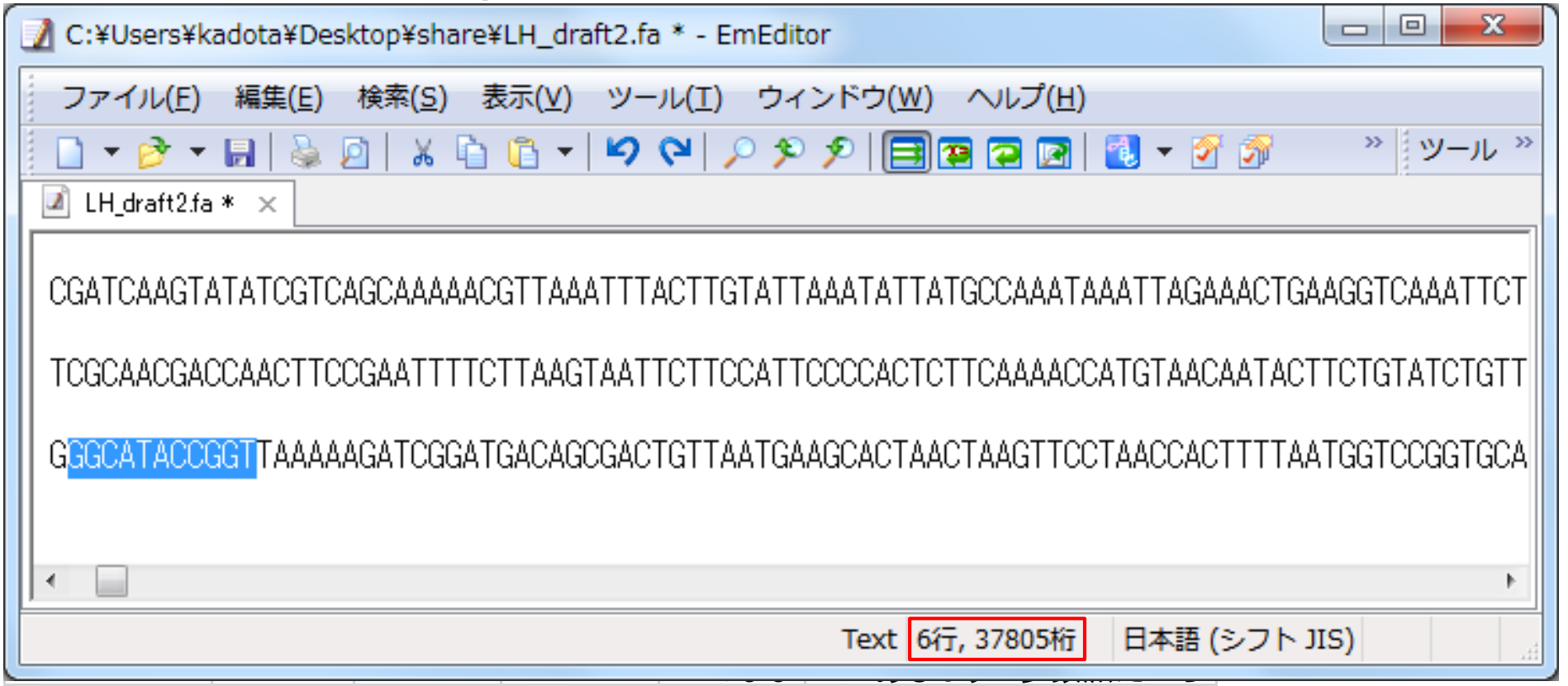


plasmid1	66668	C	T	Cのまま	Tのほうがちょっと多いが微妙
plasmid2	2569	CAAAAA	CAAAA	Aを欠失	そのほうが多数派だから
plasmid2	37805	TAAAAA	TAAAA	Aを欠失	そのほうが多数派だから

①

W27-3:37,805番目

① 37794-37804番目の11塩基分
(GGCATACCGGT)での検索結果

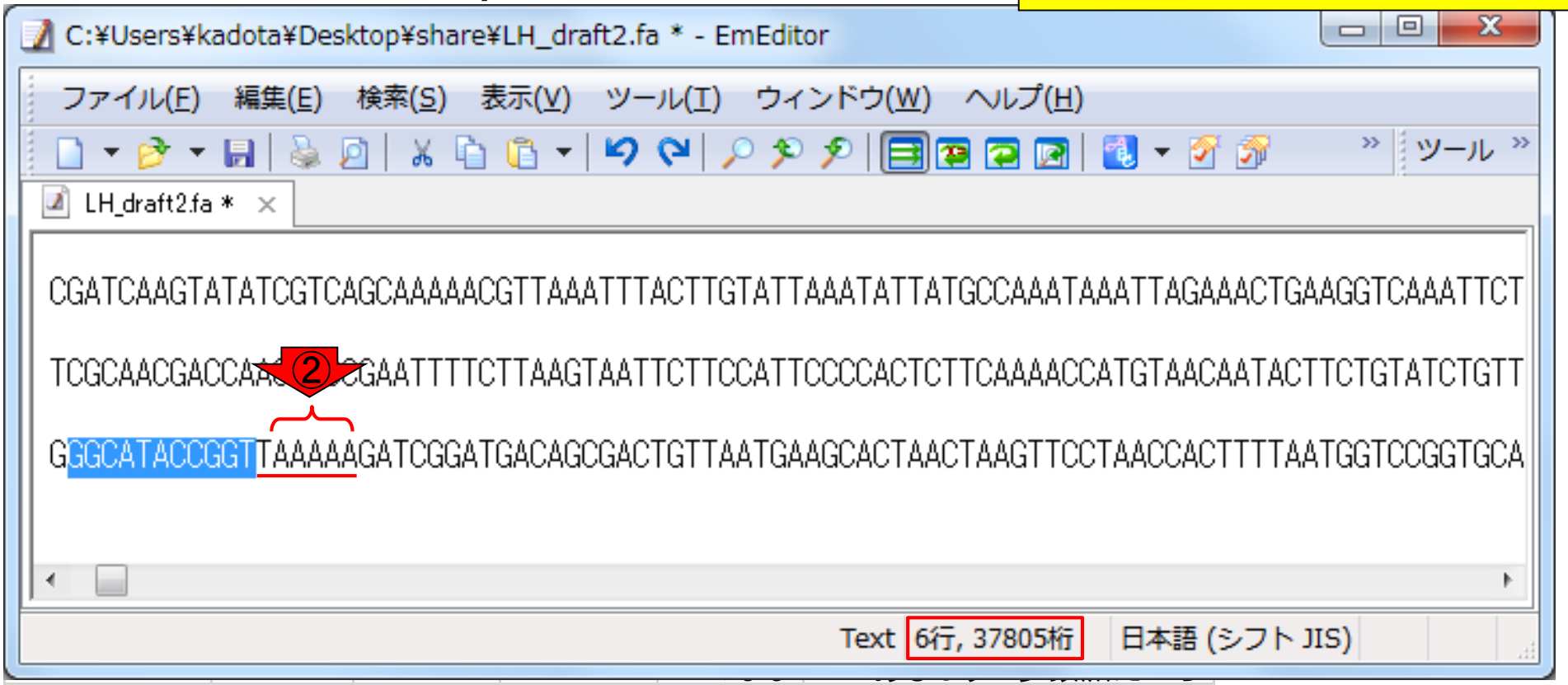


chromosome	2056279	A	G	Aのまま	Aが多数派だから
chromosome	2057564	T	C	Tのまま	Tが多数派だから
chromosome	2058145	C	T,G	Tに変更	Tが多数派だから
plasmid1	37177	G	T	Gのまま	Gが多数派だから
plasmid1	66609	A	G	Aのまま	Gのほうがちょっと多いが微妙
plasmid1	66637	C	T	Cのまま	Tのほうがちょっと多いが微妙
plasmid1	66668	C	T	Cのまま	Tのほうがちょっと多いが微妙
plasmid2	2569	CAAAAA	CAAAA	Aを欠失	そのほうが多数派だから
plasmid2	37805	TAAAAA	TAAAA	Aを欠失	そのほうが多数派だから

①

W27-3：37,805番目

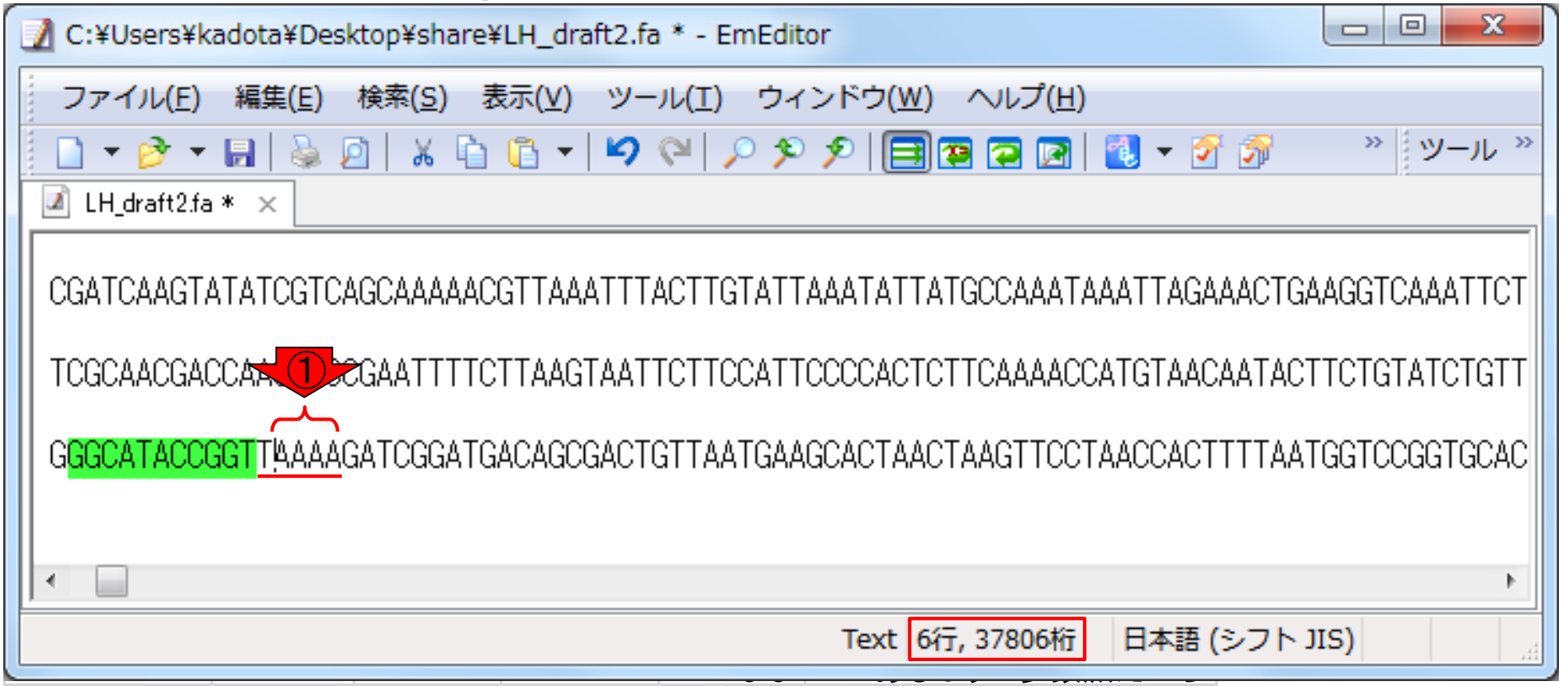
実際に変更するのは①37,805番目のTではなく、②37,806-37,810番目の中のいずれか1つのAである点に注意！



chromosome	2056279	A	G	Aのまま	Aが多数派だから
chromosome	2057564	T	C	Tのまま	Tが多数派だから
chromosome	2058145	C	T,G	Tに変更	Tが多数派だから
plasmid1	37177	G	T	Gのまま	Gが多数派だから
plasmid1	66609	A	G	Aのまま	Gのほうがちょっと多いが微妙
plasmid1	66637	C	T	Cのまま	Tのほうがちょっと多いが微妙
plasmid1	66668	C	T	Cのまま	Tのほうがちょっと多いが微妙
plasmid2	2569	CAAAAA	CAAAA	Aを欠失	そのほうが多数派だから
plasmid2	① 37805	TAAAAA	TAAAA	Aを欠失	そのほうが多数派だから

①Aを1つ削除した後の状態

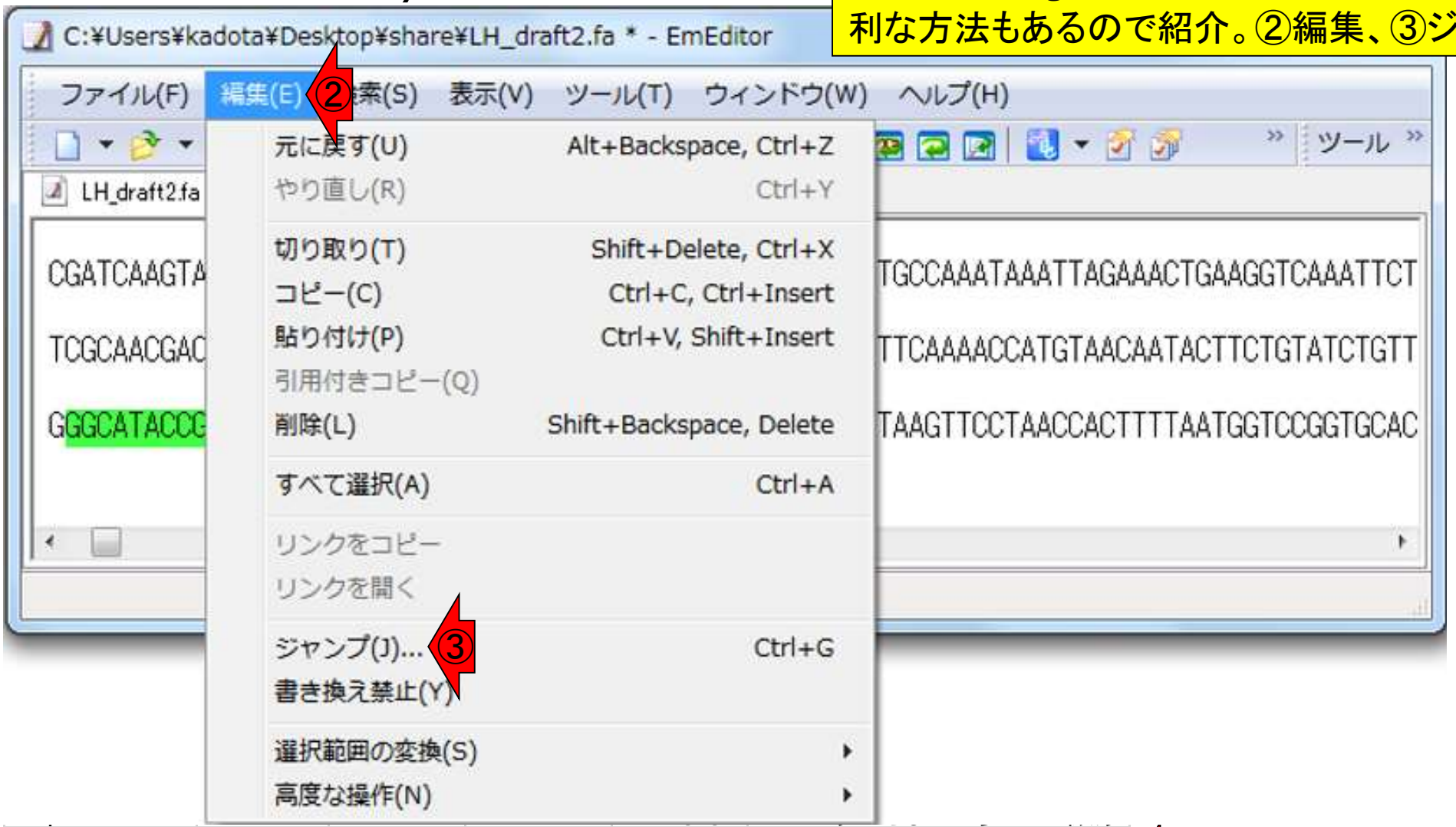
W27-3：37,805番目



chromosome	2056279	A	G	Aのまま	Aが多数派だから
chromosome	2057564	T	C	Tのまま	Tが多数派だから
chromosome	2058145	C	T,G	Tに変更	Tが多数派だから
plasmid1	37177	G	T	Gのまま	Gが多数派だから
plasmid1	66609	A	G	Aのまま	Gのほうがちょっと多いが微妙
plasmid1	66637	C	T	Cのまま	Tのほうがちょっと多いが微妙
plasmid1	66668	C	T	Cのまま	Tのほうがちょっと多いが微妙
plasmid2	2569	CAAAAA	CAAAA	Aを欠失	そのほうが多数派だから
plasmid2	37805	TAAAAA	TAAAA	Aを欠失	そのほうが多数派だから

W27-4：2,569番目

①次は2,569番目。配列数（コンティグ数）が多い場合には意味をなさないが、今回のような3配列（3 contigs）しかない場合には、非常に便利な方法もあるので紹介。②編集、③ジャンプ

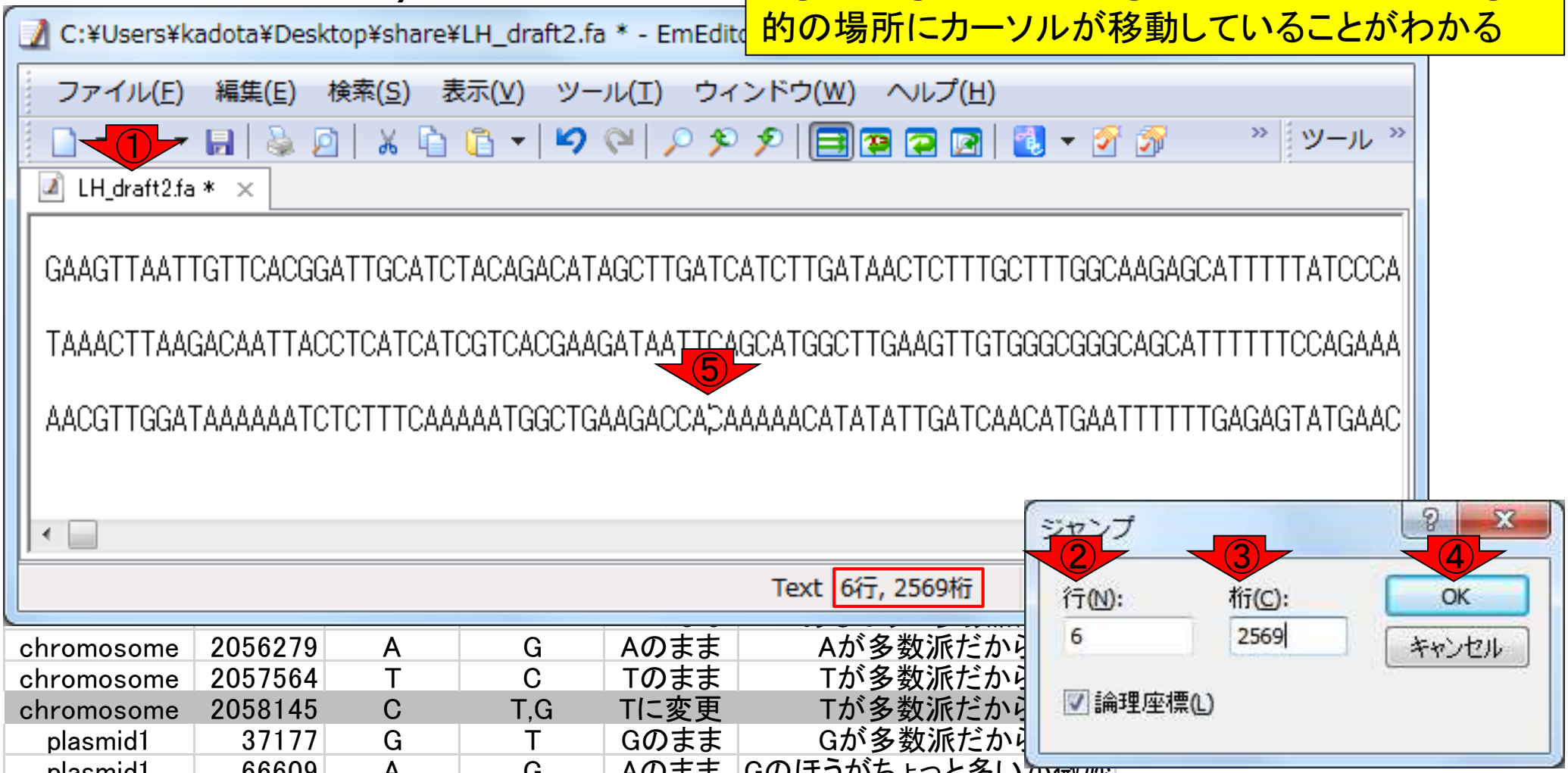


plasmid1	66668	C	T	Cのまま	Tのほうがちょっと多いが微妙
plasmid2	2569	CAAAAA	CAAAA	Aを欠失	そのほうが多数派だから
plasmid2	37805	TAAAAA	TAAAA	Aを欠失	そのほうが多数派だから

①

W27-4：2,569番目

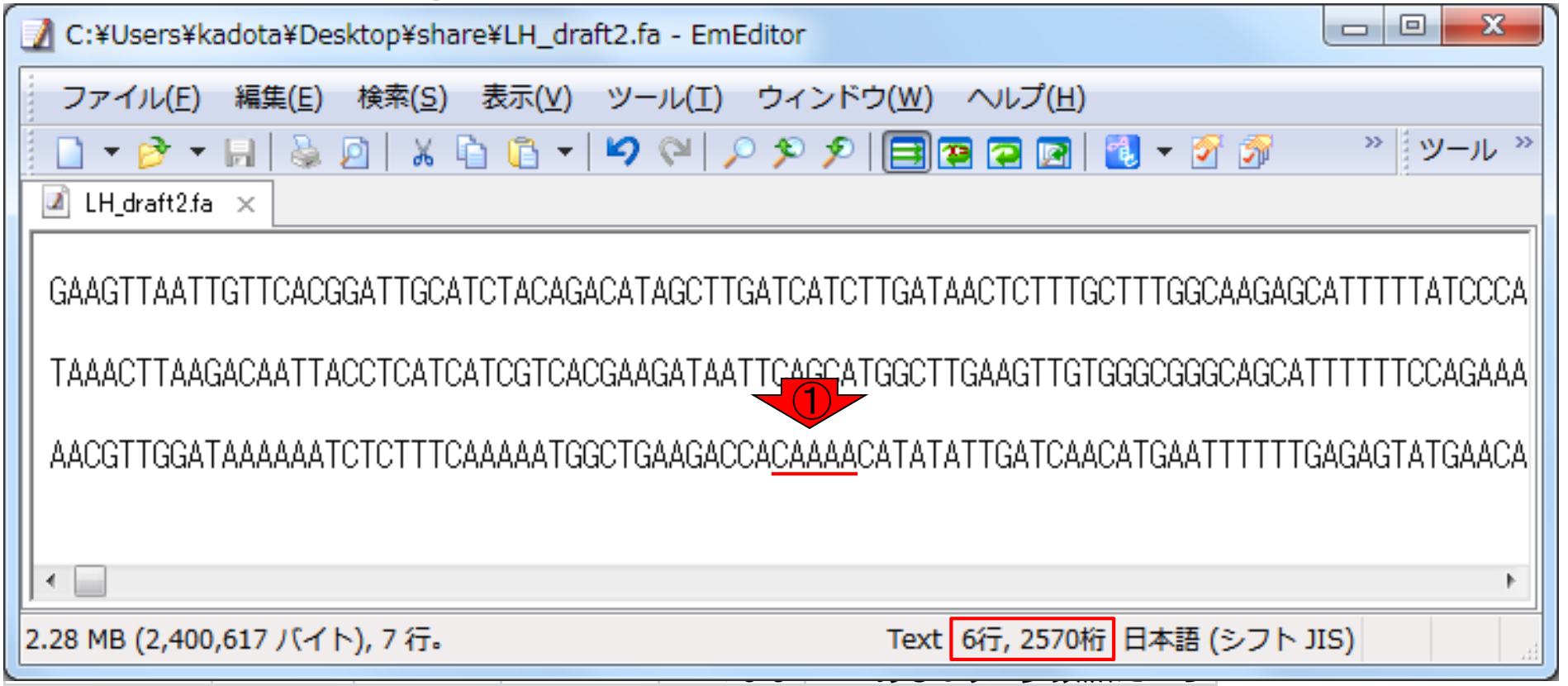
今編集しているのは、①multi-FASTA形式のLH_draft2.fa。これの3番目の配列がplasmid2なので、②6行、③2569桁目に④ジャンプすればよい。⑤目的の場所にカーソルが移動していることがわかる



chromosome	2056279	A	G	Aのまま	Aが多数派だから
chromosome	2057564	T	C	Tのまま	Tが多数派だから
chromosome	2058145	C	T,G	Tに変更	Tが多数派だから
plasmid1	37177	G	T	Gのまま	Gが多数派だから
plasmid1	66609	A	G	Aのまま	Gのほうがちょっと多いが微妙
plasmid1	66637	C	T	Cのまま	Tのほうがちょっと多いが微妙
plasmid1	66668	C	T	Cのまま	Tのほうがちょっと多いが微妙
plasmid2	2569	CAAAAA	CAAAA	Aを欠失	そのほうが多数派だから
plasmid2	37805	TAAAAA	TAAAA	Aを欠失	そのほうが多数派だから

①Aを1つ削除した後の状態

W27-4：2,569番目



chromosome	2056279	A	G	Aのまま	Aが多数派だから
chromosome	2057564	T	C	Tのまま	Tが多数派だから
chromosome	2058145	C	T,G	Tに変更	Tが多数派だから
plasmid1	37177	G	T	Gのまま	Gが多数派だから
plasmid1	66609	A	G	Aのまま	Gのほうがちょっと多いが微妙
plasmid1	66637	C	T	Cのまま	Tのほうがちょっと多いが微妙
plasmid1	66668	C	T	Cのまま	Tのほうがちょっと多いが微妙
plasmid2	2569	CAAAAA	CAAAA	Aを欠失	そのほうが多数派だから
plasmid2	37805	TAAAAA	TAAAA	Aを欠失	そのほうが多数派だから

W27-5：残りの2箇所も

同じノリで①、②の順番で数値の大きいほうから反映させておく。位置を表す数値が大きいものから順番に行えば、位置のずれに悩まされることはないため

#CHROM	POS	REF	ALT	方針	理由
chromosome	141598	A	G	Aのまま	Aが5個、Gが2個だったから
chromosome	242819	T	C	Tのまま	Tが7個、Cが2個だったから
chromosome	359141	T	C	Tのまま	Tが多数派だから
chromosome	463882	T	G	Tのまま	Tが2個、Gが1個だったから
chromosome	660031	C	T	Cのまま	Cが8個、Tが2個だったから
chromosome	792963	T	C	Tのまま	Tが多数派だから
chromosome	887422	G	A	Gのまま	Gが多数派だから
chromosome	913470	T	C	Tのまま	Tが多数派だから
chromosome	965171	T	C	Tのまま	Tが多数派だから
chromosome	1134950	A	G	Aのまま	Aが多数派だから
chromosome	1160048	A	G	Aのまま	Aが多数派だから
chromosome	1160095	A	G	Aのまま	Aが多数派だから
chromosome	1240248	GCCCC	GCCCCC	Cを挿入	そのほうが多数派だから
chromosome	1455552	CAAAA	CAAAAA	Aを挿入	そのほうが多数派だから
chromosome	1475721	T	C	Tのまま	Tが多数派だから
chromosome	1752680	G	A	Gのまま	Gが多数派だから
chromosome	1941380	T	C	Tのまま	Tが多数派だから
chromosome	2052955	A	G	Aのまま	Aが多数派だから
chromosome	2054084	TC	T	TCのまま	Cがあるほうが多数派だから
chromosome	2056279	A	G	Aのまま	Aが多数派だから
chromosome	2057564	T	C	Tのまま	Tが多数派だから
chromosome	2058145	C	T,G	Tに変更	Tが多数派だから
plasmid1	37177	G	T	Gのまま	Gが多数派だから
plasmid1	66609	A	G	Aのまま	Gのほうがちょっと多いが微妙
plasmid1	66637	C	T	Cのまま	Tのほうがちょっと多いが微妙
plasmid1	66668	C	T	Cのまま	Tのほうがちょっと多いが微妙
plasmid2	2569	CAAAAA	CAAAA	Aを欠失	そのほうが多数派だから
plasmid2	37805	TAAAAA	TAAAA	Aを欠失	そのほうが多数派だから

②

①

W27-5：残りの2箇所も

同じノリで①、②の順番で数値の大きいほうから反映させておく。位置を表す数値が大きいものから順番に行えば、位置のずれに悩まされることはないため

#CHROM	POS	REF	ALT	方針	理
chromosome	141598	A	G	Aのまま	Aが5個、Gが2
chromosome	242819	T	C	Tのまま	Tが7個、Cが2個だったから
				Tのまま	Tが多数派だから
				Tのまま	Tが2個、Gが1個だったから
				Cのまま	Cが8個、Tが2個だったから
				Tのまま	Tが多数派だから
				Gのまま	Gが多数派だから
				Tのまま	Tが多数派だから
				Tのまま	Tが多数派だから
				Aのまま	Aが多数派だから
chromosome	1160048	A	G	Aのまま	Aが多数派だから
chromosome	1160095	A	G	Aのまま	Aが多数派だから
chromosome	1240248	GCCCC	GCCCCC	Cを挿入	そのほうが多数派だから
chromosome	1455552	CAAAA	CAAAAA	Aを挿入	そのほうが多数派だから
chromosome	1475721	T	C	Tのまま	Tが多数派だから
chromosome	1752680	G	A	Gのまま	Gが多数派だから
chromosome	1941380	T	C	Tのまま	Tが多数派だ
chromosome	2052955	A	G	Aのまま	Aが多数派だ
chromosome	2054084	TC	T	TCのまま	Cがあるほうが多数
chromosome	2056279	A	G	Aのまま	Aが多数派だ
chromosome	2057564	T	C	Tのまま	Tが多数派だ
chromosome	2058145	C	T,G	Tに変更	Tが多数派だ
plasmid1	37177	G	T	Gのまま	Gが多数派だ
plasmid1	66609	A	G	Aのまま	Gのほうがちょっと多
plasmid1	66637	C	T	Cのまま	Tのほうがちょっと多いが微妙
plasmid1	66668	C	T	Cのまま	Tのほうがちょっと多いが微妙
plasmid2	2569	CAAAAA	CAAAA	Aを欠失	そのほうが多数派だから
plasmid2	37805	TAAAAA	TAAAA	Aを欠失	そのほうが多数派だから